

Following through on Phillips

Britain's inquiry into the BSE crisis has revealed significant weaknesses in the way the government used scientific advice and established research priorities on a topic of urgent social concern. These require more detailed public scrutiny.

Did Lord Rothschild get it wrong? Academic research financed by UK government departments is still based on the 'customer/contractor' relationship proposed by Rothschild almost 30 years ago. Last week, as Lord Phillips' public inquiry into the recent crisis over bovine spongiform encephalopathy (BSE) exposed the blurring of scientific and political judgements inside the Ministry of Agriculture, Fisheries and Food (MAFF), some observers were asking if Rothschild made a mistake in inviting just such an overlap.

There is certainly much in the inquiry's report to support this view. At its heart is the finding that public statements by MAFF, and its decisions on research directions, were dominated by a political objective: the need to provide reassurances about the safety of British beef. Evidence given to the Phillips committee reveals the wide impact of this mind-set. For example, the political goal of reassurance seems to have been closely linked to support for the scientific hypothesis that BSE was a variant of scrapie, present in UK sheep for many centuries.

Such evidence might be taken as a warning against ever allowing political agendas to influence scientific priorities. But a different conclusion can also be drawn, that the problem was not the attempt to impose policy-related priorities on research in an area of vital social concern, but rather the misguided nature of those priorities.

In other circumstances, political priorities have shaped the research agenda in a more positive way. The UK government's decision to appoint an AIDS research supremo in the 1980s to review the

scientific agenda, compare it with public-health needs, and commission research to fill perceived gaps, was one example. If the BSE research agenda had been controlled by the Department of Health rather than MAFF, with public rather than animal health dominating the agenda, the outcome might have been very different.

In other words, Rothschild got it right, at least up to a point. The failure was one of implementation, rooted in political dynamics. The influence of political muscle is highlighted by the ill-fated attempts by health-department officials in 1990 to persuade MAFF to appoint a single research supremo for BSE, similar to the AIDS post (see page 5). A striking feature of Phillips' description of this episode is the absence within government of any individual with sufficient clout to establish a coherent research policy on BSE, properly informed by the nation's leading scientific experts.

Why this situation arose is a question of pressing national interest, sufficiently pressing to justify a separate inquiry into the linked dynamics of science advice and priority setting. The Royal Society has taken a step in the right direction by promising a quick report on the governance of science. But the Phillips report provides a unique opportunity to go further. A new public inquiry could revisit sensitive topics, from the quality of research in government labs to the ways in which scientific advice is put into practice. It should also address whether BSE research should be handled by the Food Standards Agency, not MAFF. The challenge is daunting, but the need is urgent. ■

Total eclipse unlikely

Even as large optical telescopes steal much of the limelight, smaller instruments can retain an important role.

The world of astronomy is entering a new era of large optical telescopes — giant pieces of equipment with primary mirrors measuring eight or more metres across. The light-gathering capacity of these enormous 'photon buckets' gives them inherent advantages: astronomers can observe fainter sources, and therefore address problems that were beyond the capacity of smaller telescopes. They can also observe brighter sources much more quickly, speeding up project execution.

Add to that the fine resolution made possible by combining a large mirror with adaptive-optics techniques, which correct for the distorting effects of atmospheric turbulence, and it is easy to see why the operators and users of small telescopes feel threatened. Money for astronomy comes from a limited pot, and with titanic telescopes consuming an ever-bigger share of it, many smaller observatories are coming under pressure to close (see page 12).

But although some older telescopes will have to go, it would be a mistake to cull the majority. For a start, many projects don't require giant mirrors, and can be efficiently conducted using a modest-sized telescope. And any study requiring repeated observations of the same object can only realistically be done using a small telescope. If the discoverers of extrasolar planets had had to compete for observing slots on one of the 10-metre Keck telescopes, for example, their high-risk endeavour would never have been granted enough time to bear fruit.

There is also a social factor: the shift towards fewer, larger telescopes inevitably places power in the hands of a privileged élite of astronomers. If diversity is strength, then, for the good of the astronomical community, it is important to retain a sufficient number of smaller telescopes to allow independent groups to do their own thing.

But given limited budgets, the users of small telescopes will have to look beyond a 'business as usual' approach. The small telescopes that survive into the era of the giants will be those that best demonstrate their efficiency and scientific productivity. That may mean that telescopes will have to concentrate on focused projects, rather than holding open competitions for observing time. It may also mean that ownership of these telescopes should pass from centralized organizations with relatively high overhead costs to leaner and fitter consortia of universities or research groups.

Finally, astronomers will need to change their attitudes towards the technical specialists who build instrumentation. If small telescopes are to work on focused projects, they will need to be fitted with highly specific instruments, optimized for the task in hand. Currently, those who build detectors and other instruments are sometimes treated as second-class citizens by certain of their colleagues. Their skills should be granted the respect already enjoyed by the builders of instruments in other scientific disciplines such as space science and high-energy physics. ■





Trade secret
Kenyan discovery has led to a potential HIV vaccine
p6



Join up
Report says fusion scientists are too isolated
p7



Fishy practice
Certification plan to end use of cyanide in aquarium fish trade
p8



Wing of bat...
Boat designers get in flap over a new jointed mast
p10

Inquiry blames missed warnings for scale of Britain's BSE crisis



Peter Aldhous, London

"Beef is safe." That supposedly comforting phrase became a mantra for the British government after bovine spongiform encephalopathy (BSE) emerged in the late 1980s. No wonder the public felt betrayed when on 20 March 1996 they were told by the same government that a handful of deaths from a new variant of Creutzfeldt-Jakob disease (vCJD), an invariably fatal neurodegenerative condition, were probably linked to BSE.

The long-awaited report from the public inquiry into the official handling of Britain's BSE epidemic, released last week, concludes that ministers and civil servants did not deliberately lie to the public — they genuinely believed that the risks were minimal.

But it identifies institutional deficiencies which led to a mistaken campaign of reassurance that failed to reflect the shifting state of scientific knowledge. "Most of those responsible for responding to the challenge posed by BSE emerge with credit," says the report, from a team headed by the judge Lord Phillips. "However, there were a number of shortcomings in the way things were done."

The report's lessons should reverberate internationally. And it has already been delivered to the governments of other Euro-

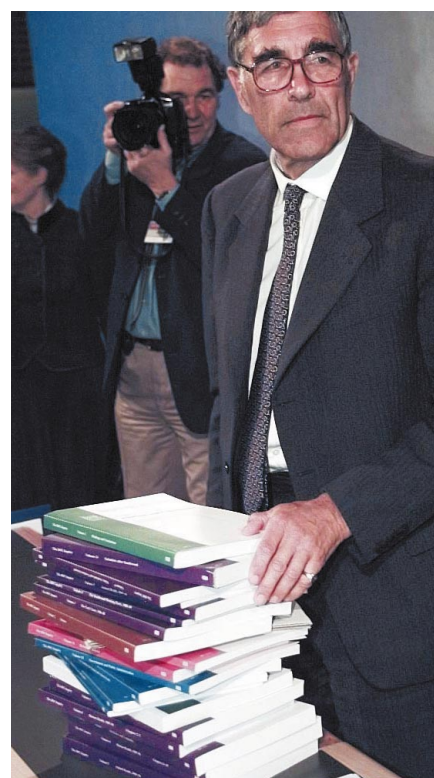
Phillips: says that British ministers believed the risks to the public were minimal.

pean Union states — some of which are now wrestling with their own emerging BSE outbreaks, linked to imported British cattle feed.

The inquiry's report documents failings in the scientific advisory system that delayed measures that might have reduced the size of Britain's BSE epidemic, or limited the amount of potentially infective material entering the human food chain. It addresses deficiencies in the organization of BSE research (see page 5). And it illustrates how the conclusion that BSE was likely to pose only a "remote" risk to human health, reached by a scientific advisory panel in February 1989, was allowed to dominate official thinking right up until the fateful 1996 announcement — even in the face of growing scientific evidence that should have raised warning signs to the contrary.

BSE, CJD and scrapie, a related disease of sheep, are mysterious conditions that are now widely thought to be caused by prions. These are misshapen variants of a protein called PrP. The prions seem to cause normal versions of the protein to adopt their own, distorted configuration. This results in the formation of insoluble clumps of protein in the brain, which seem to kill neighbouring cells.

When BSE emerged, the prion theory



had yet to win widespread acceptance. But the first scientific committee to consider the issue, headed by Richard Southwood, a

Were some CJD victims infected by vaccines?



Bitter pill: some oral polio vaccines have been made from potentially infected bovine material.



Karen Birmingham, London

Although food has been the overwhelming object of attention as the source of infection with variant Creutzfeldt-Jakob disease (vCJD), there is a possibility that some of the victims — who are predominantly young adults — might have contracted the disease through immunization.

The BSE inquiry's report (see above) calls for vaccines to be investigated as a possible route of transmission. But it concedes that this will be hampered by the

fact that "systematic records of the action taken in response to BSE in respect of individual medical products are lacking".

Indeed, the report reveals that the pharmaceuticals industry and related government agencies failed to grasp the extent of the risk, and delayed removing potentially harmful products from the public arena. It notes that this behaviour persisted longer than corresponding problems in the meat industry, without attracting similar parliamentary and public scrutiny.

Some vaccines are made using culture media derived from cattle tissue. Although

▶ zoologist at the University of Oxford, drew comfort from the parallels with scrapie. This disease had shown no evidence of posing a risk to human health.

Southwood's working party was told by John Wilesmith, then the sole qualified veterinary epidemiologist within the Ministry of Agriculture, Fisheries and Food (MAFF), that scrapie-infected sheep were the most likely source of BSE infection. Wilesmith correctly identified meat and bone-meal in cattle feed as the prime culprits. He noted that the first recorded cases of BSE appeared several years after the processes for rendering animal carcasses into feed were changed. He concluded that this had allowed the scrapie agent to remain active and infect cattle.

Early advice

At their first meeting in June 1988, Southwood and his colleagues were shocked to learn that cattle known to have BSE were still entering the human food chain. They advised that this should be stopped. According to Phillips, the government should have taken this action earlier. "Reference to outside committees involves delay," says his report. "It should be avoided, where possible, in situations of urgency." Indeed, the report says that, in general, the government relied too heavily on outside experts, and should build up its in-house scientific expertise.

In addition to Southwood, the working party comprised three other eminent scientists: a virologist, a neurologist and a veterinarian. This typified the practice — then common in Britain, but not uniquely British — of taking advice from a small committee of scientists known personally to government officials, and not necessarily experts in the narrow field under scrutiny. In future, the Phillips report concludes, committees "should include experts in the areas of advice that is likely to be required".

The Phillips report also identifies specific deficiencies in the advice delivered by Southwood in February 1989 — the working party



The wait is over: relatives of those infected by vCJD attended the release of the BSE inquiry's report.

is among those explicitly criticized. Although the group acknowledged that, if its assumptions were incorrect, the implications would be extremely serious, Phillips faults the working party for not making clear that it had not seen the data on which Wilesmith's epidemiological conclusions were based.

Those data were restricted by MAFF to Wilesmith's team until after the vCJD crisis broke in 1996. Only then was an independent team, led by Roy Anderson of the Wellcome Trust Centre for the Epidemiology of Infectious Disease in Oxford, allowed access. Anderson's initial studies (see *Nature* 382, 779–788; 1996) and subsequent work have cast doubt on some of Wilesmith's conclusions. Anderson now says that cows were probably infected before the changes in rendering practice were introduced, which makes the original scrapie theory less convincing.

Dangerous assumptions

Whatever the origin of BSE, it soon became clear that the infectious agent at loose in British cattle was different from known strains of scrapie. So, even if it had originally passed from sheep, BSE could no longer

be assumed to be equally benign. In 1990, a scare erupted when domestic cats started to go down with a BSE-like disease. The significance was played down by MAFF officials, but scientists were worried because it had never been possible to transmit scrapie to cats. By 1992, a team led by Moira Bruce at the Institute of Animal Health's Neuro-pathogenesis Unit in Edinburgh had found that BSE was distinct from known isolates of scrapie. And by 1994, she had proved that the cats were infected with BSE.

This evidence was reviewed by the Spongiform Encephalopathy Advisory Committee (SEAC), which advised the government on the BSE epidemic from April 1990. Acknowledging that BSE was not the same as conventional scrapie, SEAC still judged the risk to human health to be remote because tissues known to be potentially infective — the central nervous system and various other bits of offal — were supposed to be removed from the human food chain.

The trouble, according to Phillips, was that officials and ministers were still falling back on Southwood's conclusion of a remote risk to human health. "Right up to 1996 the

▶ guidelines asking drugs companies to get this material from countries with BSE-free herds were issued in 1989 by the Committee on Safety of Medicines (CSM), they will not become law until March 2001. Companies have been relied on to follow the guidelines voluntarily. Yet only last month, it emerged that one company, Medeva, had continued to use bovine serum from British herds to make an oral polio vaccine (see *Nature* 407, 936; 2000). The Department of Health delayed release of this news for two months.

Alarming, vaccines produced after the point at which the BSE epidemic had been identified — possibly using British bovine material — were still in use as recently as November 1993. According to the inquiry's

report, the chief medical officer of the day, Donald Acheson, decided to phase out the existing stocks because new batches of vaccines take time to grow, and medical experts considered that the benefit of maintaining a continuous national immunization programme outweighed the risk of interrupting it.

These decisions were taken in the absence of knowledge about how infective bovine materials used to make vaccines might be. It was only in 1993 that the Neuropathogenesis Unit in Edinburgh obtained test results that showed bovine serum did not cause infection in mice — studies that the inquiry's report describes as "inconclusive".

But in some sections, the report is

sympathetic to the difficulties of withdrawing medicines from use quickly. Product licences would have had to have been revoked on an individual basis, so each licence holder could have appeared before the relevant committee to argue against withdrawal. "This would have been a logistical nightmare," says the report. "It is not surprising that the decision was taken to issue guidelines rather than attempt to use formal statutory powers."

Last week it was also revealed that the Department of Health asked the CSM as recently as March this year to produce a comprehensive assessment of BSE-related issues in vaccines. There has been no indication when this evaluation will be made public.

Scientific turf war 'delayed potentially vital research'



David Dickson, London

Individual scientists may have largely escaped criticism in Britain's BSE inquiry, chaired by Lord Phillips. But the inquiry's report sharply highlights critical weaknesses in the way research decisions were made.

In particular, the report suggests that turf fights between the research councils and government departments may have delayed vital research in the early 1990s. Such tensions, it reveals, undermined efforts in 1990 by Donald Acheson, then chief medical officer at the Department of Health, to give responsibility for BSE-related research to a single research supremo. Scientists looking at the implications of the BSE crisis are likely to focus on the factors that led to this failure.

Acheson wanted a single individual to be responsible for identifying key research needs and commissioning the appropriate additional research in a way that had already been successfully applied to the AIDS epidemic. The Phillips report identifies various areas — such as experimental testing of the 'scrapie' hypothesis (see main story) and the development of a reliable diagnostic blood test — where progress might have been more rapid if a supremo had been in place.

But, as evidence submitted to the inquiry shows, the proposal fell victim to a basic distrust between government officials and scientists working in universities or for the research councils. In particular, officials from the Ministry of Agriculture, Fisheries and Food (MAFF) remained adamant that any attempt to develop and monitor a broad research agenda should be under their control — and be directed by a veterinary expert rather than a scientist with a background in fundamental research.

At the same time, according to evidence presented to the inquiry, representatives of the two research councils directly involved — the Medical Research Council and the Agricultural and Food Research Council — were keen to remain independent from anything that felt like government interference.

These differences could not be reconciled. And the minutes of a subsequent meeting of government officials records a general feeling that the supremo idea "should be left on one side to die a natural death".

Apart from one area of research — the possibility of maternal transmission of BSE from cows to their calves — the Phillips report rejects the charge that BSE research suffered from inadequate funding.

But it does draw attention to the lack of collaboration between scientists working in

government laboratories and non-government researchers. For example, the report points out that scientists from MAFF's Central Veterinary Laboratory (CVL) in Weybridge, Surrey, held back initially from consulting scientists at the Neuropathogenesis Unit (NPU) in Edinburgh. "If the CVL had consulted them at this stage, the NPU would have confirmed that there were very strong indications that this was indeed a new Transmissible Spongiform Encephalopathy," it says.

The evidence unearthed by the Phillips inquiry has provided a rich source of ammunition for those keen to see a reassessment of the division of responsibilities between government and non-government scientists on topics of public concern.

Peter Cotgreave, for example, director of the lobby group Save British Science, says that the one clear lesson is "recognition of the need for the scientific community to get more closely involved" in deciding on research priorities in areas with important policy implications.

Eric Millstone, of the Science Policy Research Unit at the University of Sussex, agrees — but says that this does not necessarily mean passing responsibility for choosing priorities to bodies such as the research councils. "Phillips talks as if scientific research could and should be conducted in a policy vacuum," says Millstone. "But that is unrealistic. The failure of BSE was not that BSE research was being overseen by a government department, but that it was the wrong department."

The Royal Society has already announced plans to carry out a rapid review of the lessons of the BSE crisis for the interaction between the government and science advisory committees. Reacting with unprecedented speed, it has promised that its conclusions will be completed in December.



Southwood Report was cited as if it demonstrated a matter of scientific certainty, rather than provisional opinion," the report notes. As a result, precautionary measures designed to protect both people and cattle from infection were not rigorously enforced.

Phillips also identifies specific breakdowns in communication that exacerbated the situation. For instance, many SEAC members remained unaware that mechanical recovery techniques, used to strip meat from the spinal column, were likely to allow traces of spinal cord to get into meat products. As a result, SEAC did not suggest that the practice should be banned until November 1995. And one key piece of evidence that first emerged in November 1990, showing that BSE had been passed to a breed of sheep resistant to scrapie, was never commented on by SEAC.

Protecting the future

Several principles to stop such failings are outlined by the Phillips report. It says that scientific advice should state the reasons for its conclusions, and that underlying assumptions should be made clear. Where appropriate, the advice should set out the different policy options and the implications of each.

The report also advocates greater openness, and stresses that officials and ministers must recognize when the expertise of scientists ends, and decisions have moved into the policy arena. "Scientists were expected to provide all the answers, where sometimes this was not their proper role," says Phillips.

The recommendations will not surprise British scientists, as almost all of them have already been introduced by Robert May, the government's chief scientific adviser until two months ago.

What will surprise some scientists are the statements about BSE's origins made in the Phillips report. It says that the disease "probably originated from a novel source in the early 1970s, possibly a cow or other animal that developed disease as a consequence of gene mutation". Some feel that this goes beyond the available evidence. "It is likely to be impossible to know," says Anderson, who will soon take up a post at Imperial College, London.

The current government seems to feel this point needs clarification, and has said it intends to commission an independent review of scientific understanding of the origins of the BSE epidemic. Researchers in the field hope that this does not deflect attention from the most urgent priorities, which include developing a test to identify animals and people incubating the disease, and drugs to treat vCJD victims.

http://www.bseinquiry.gov.uk

Kenyans protest at being left off AIDS patent

Wachira Kigotho, Nairobi

The terms of a collaboration on a potential HIV vaccine, between a British research team and the University of Nairobi, are to be revised after the Kenyan scientists complained that their names had been excluded from a patent application on the vaccine.

The original patent on the subtype A DNA candidate vaccine was filed last year by the UK Medical Research Council in the names of Andrew McMichael and Tomas Hanke, of Oxford University's Institute of Molecular Medicine.

The vaccine is based on a mechanism by which certain HIV-resistant prostitutes in Nairobi appear to be immune to the disease. McMichael and Hanke have been primarily responsible for identifying the molecular basis of this mechanism and building it into the vaccine.

Under an agreement reached with the University of Nairobi two weeks ago, a new memorandum will be drawn up recognizing the contribution of scientists in the university's Department of Microbiology. These researchers have been recording the T-cell responses among prostitutes in the Majengo slums of Nairobi for more than a decade.

The dispute arose from the fact that, although the two teams work closely together — trials of the vaccine are already under way in Oxford, and are due to start in Kenya shortly — the Oxford researchers had not told their African colleagues that they were applying for the patent.

"As soon as we realized our names were not included, we entered into correspondence with our partners to ensure that we were reflected as being among the researchers," says Job Bwayo, head of microbiology and team leader of the AIDS vaccine initiative programme at the University of Nairobi.

According to McMichael, his research team took out the patent primarily to prevent other organizations from doing so. He says that his and Hanke's names are the only ones on the application because its novel technical content — primarily the sequence of the construct — was based entirely on work carried out in Oxford.

"There has been no intention of shutting out the Kenyans from the fruits of the research," says McMichael, although he admits that "we should have raised the patent issue much earlier".

He adds that one aim of seeking a patent was to ensure that, if the clinical trials prove successful, the vaccine would not be marketed at a profit, but at a price that most Africans would be able to afford.

It has now been agreed that the original memorandum of understanding will be modified to recognize the contribution of



Prostitutes in Nairobi wait for a medical check-up. A potential HIV vaccine has been developed from Kenyan prostitutes immune to the virus.

Bwayo and his colleagues Omu Anzala and Jeckoniah Ndinya-Achola, who have been involved in the broader research behind the potential vaccine.

Talks chaired by the vice-chancellor of the University of Nairobi, Francis Gichaga, have also led to an agreement to set up a task force consisting of patent and intellectual-property experts from the two universities and members of the New York-based International Aids Vaccine Initiative (IAVI), the main sponsor of the clinical trials.

One issue still to be addressed is how to

benefit the 60 prostitutes whose immune systems provided the knowledge to develop the vaccine. "We have not forgotten them," says Anzala, the project's programme manager in Nairobi.

A statement issued by the University of Nairobi, the UK Medical Research Council and IAVI, while acknowledging "shortfalls" in the current memorandum, also stresses that "this patent was filed in good faith to protect the candidate DNA vaccine from unauthorized third-party exploitation".

♦ <http://www.iavi.org>

Puffer fish joins genome stampede

David Cyranoski, Tokyo
& Paul Smaglik, Washington

An international consortium led by the US-based Joint Genome Institute last week announced plans to sequence the genome of the puffer fish *Fugu rubripes*. It is hoped that this will offer a cheap and fast route to unravelling many of the mysteries of the human genome.

Since the draft sequence of the human genome was announced in June, annotation of the sequence — allocating genes to pieces of sequence — has progressed relatively slowly. Comparative analysis with other vertebrate genomes is thought to be the key to illuminating many of these genes and regulatory sequences, and another consortium has already announced that it will produce a draft of the mouse genome by next March (see *Nature* 407, 663–664; 2000).

The *Fugu* genome, also expected to be ready in draft form by next March or April, should make finding the genes a much faster process than if only the human and mouse genomes were to hand. "It's going to have a pretty dramatic effect," says Trevor Hawkins, deputy director of the Joint



Economy size: *Fugu* has as many genes as humans and mice, but a much smaller genome.

Genome Institute (JGI). "I can't quantify it, but it's going to be pretty amazing."

The JGI, which is based at the Lawrence Berkeley, Lawrence Livermore and Los Alamos national laboratories, will use a procedure called the whole-genome shotgun technique. This breaks the genome into fragments, after which 84 automated sequencers will analyse the 400 million base pairs of DNA in the *Fugu* genome. Raw sequence data will be released every 24 hours, in accordance with the public project's policy. Scientists can then use these gene fragments to run searches against genes or potential genes in other organisms.

By March, the JGI should have sequenced the fish's genome at least three times over, covering 95% of the genome. This should provide enough data to assemble the fragments into contiguous stretches of DNA, and allow comparisons with the mouse and human genomes.

"I want to take it higher," says Hawkins. Higher coverage, perhaps to four or five times, would result in longer stretches of unbroken DNA, and therefore better comparisons.

Other members of the consortium include Singapore's Institute of Molecular and Cell Biology (IMCB), the US Institute of Systems Biology in Seattle, the UK Human Genome Mapping Project Resource Centre, and the US Molecular Sciences Institute in Berkeley, California. They will share the computational load and be responsible for 'finishing' the *Fugu* genome by putting together the fragments produced by the JGI's shotgun technique. The finishing phase is expected to be complete by the spring of 2002.

The *Fugu* genome is thought to have about the same number of genes as the human genome — 60,000 to 65,000, according to estimates by IMCB's B. Venkatesh. But the *Fugu* genes are spread over only 400 million DNA base pairs, compared with some 3 billion base pairs in the human and mouse. "It can be thought of as the minimum set of genes to build a vertebrate," says Venkatesh.

Comparing *Fugu* and human "is a very cheap way to get a gene catalogue", says Jean Weissenbach, director of the Centre National de Séquençage in Evry, France. This spring, Weissenbach used the partial sequence of a related puffer fish, *Tetraodon nigroviridis*, to predict that the human genome has 28,000–34,000 genes. He expects the *Fugu* sequence to confirm that estimate.

And because that genome contains few repeats or 'junk' DNA, it should be easier to find similar sequences — and hence genes — between *Fugu* and human. Human, mouse and *Fugu* genes all share exons — lengths of DNA that code for proteins. But in the two mammalian genomes, some introns — stretches of DNA that are not translated into the final protein — also come as part of the package. Having *Fugu*'s sequence will help weed out those non-coding areas, says Weissenbach.

For IMCB's Chris Tan the project is both a way for Singapore to develop in bioinformatics and biocomputation and a way to move genomics beyond sequencing to "really exciting applications and functional analysis". ■

▶ <http://jgi.doe.gov/tempweb/programs/fugu.htm>

US fusion community 'must end isolation', says panel

Colin Macilwain, Washington

The quality of fusion research in the United States is on a par with other areas of the physical sciences, according to an assessment by the National Academy of Sciences. But the fusion community is isolated from other scientists, struggles to attract young talent, and tends to put the pursuit of fusion energy ahead of the pursuit of good science, the academy warns.

The assessment, undertaken by a panel chaired by Charles Kennel, director of the Scripps Institution of Oceanography in La Jolla, California, recommends that fusion scientists make "a systematic effort to reduce [their] scientific isolation".

It also calls on the National Science Foundation to increase its involvement in fusion research — most US magnetic-fusion work is supported by a \$250-million-a-year programme at the Department of Energy (DoE).

DoE officials welcomed the report's endorsement of the quality of science in their programme. But fusion scientists — many of them gathered in Québec City in Canada last week for the American Physical Society's plasma-physics meeting — were taken aback by the harshness of the panel's judgement on the discipline, which it branded as "intellectually isolated from the rest of science".

"Most scientists funded by the program do not, in general, actively participate in the wider scientific culture," says the executive summary of the panel's report. "As a result, the flow of scientific information both out of and into the field has stagnated."

But in its main finding, the panel said that the general quality of science funded by the DoE magnetic-fusion programme was "easily on a par with other leading areas of contemporary physical science".

Four years ago the fusion programme lost



Nice work: the report praised fusion research, such as the National Spherical Torus Experiment.

one-third of its budget and the DoE changed its main goal from developing fusion-based energy sources to improving the scientific understanding of fusion. But according to Kennel, programme resources are still allocated largely to approaches that demonstrate progress towards fusion power.

Some fusion scientists argue that their experiments need to be relevant to fusion power. Magnetic fusion "is a mission-oriented science", says David Baldwin, senior vice-president of the fusion group at General Atomics in San Diego, California.

But others welcomed the report's call for more emphasis on scientific questions. Bill Dorland, a theorist at the University of Maryland's Institute for Plasma Research, says time on large fusion facilities tends to go to experiments that promise to raise technical performance, rather than understand why performance is restricted. ■

Whitehead enters into array deal

Steve Nadis, Boston

The high-tech manufacturers Corning have entered into a \$10 million research partnership with the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology. Corning will give the Whitehead \$2.5 million for each of the next four years for the development of what the company calls "the next generation of DNA microarrays and other tools of the post-genome world".

The partnership aims to produce DNA chips that can analyse 10,000 human genes

at a time and to develop protein arrays to facilitate the simultaneous study of thousands of proteins.

"There is a critical need for researchers in academia and industry to collaborate in the battle against human disease in order to develop technologies with the potential to revolutionize medical science," says Richard Young, a biologist at the Whitehead and MIT who will direct the project. "We intend to use technologies developed through the joint initiative with Corning to lay a foundation for better drugs and vaccines." ■

Aquarium group fights 'cyanide fishing'...

Mark Schroepe, Bali

The Marine Aquarium Council (MAC), an international non-profit consortium based in Honolulu, Hawaii, hopes to eliminate reef-threatening practices such as 'cyanide fishing', by launching a voluntary certification programme next year.

Tactics such as cyanide fishing — stunning fish by spraying them or the crevices in which they live with cyanide — are rampant in some sectors of the trade in live marine aquarium animals. They are used widely in key exporting countries such as Indonesia, where local governments often lack the resources to enforce bans.

The poison causes severe damage to corals and often leads to premature death of the fish. Another major problem is the over-harvesting of targeted species, which include reef fish, invertebrates (such as sea slugs, shrimps and crabs) and corals that are collected for use in aquaria.

The MAC plans to address these concerns by issuing guidelines banning the use of all chemicals in such fishing. The guidelines would be enforced through spot checks, with businesses that comply receiving certification for their products.

The incentives for following the guide-



Toxic trap: the use of cyanide to stun fish destined for aquaria is widespread in Indonesia.

lines include an enhanced environmental image, and increased demand for certified animals, as these would be healthier. Paul Holthus, executive director of the MAC, says it is unclear yet what effect the programme will have on prices. But, he says, "the marketplace is already paying for the quality when and where they can verify it".

The council is also working to begin monitoring the number of any given species imported and to increase communication

between importers and exporters. Fishermen can then be kept informed of which species are not currently needed, thus avoiding unnecessary harvesting.

At present, these data are available only for designated endangered species — which includes many corals, but no fish and few invertebrates. Such information would allow measures in response to any signs of overharvesting.

The US Coral Reef Task Force, a govern-

ment panel, has worked with the council to develop the certification system. "We're trying to build something that we both see as beneficial," says Barbara Best, a member of a task force subgroup that has drafted a proposal for new legislation for the aquarium industry.

The task force wants a separate set of certification guidelines that would probably be broadly similar to that of the MAC. Importers would have to state that their products were obtained according to the guidelines. Although the United States would not monitor collection, evidence from government or other sources abroad would lead to prosecution, with penalties for falsifying documents.

The task force and the MAC differ on some points. For instance, the proposed legislation calls for the identification of species that should not be imported because of factors such as low survival rate in transit or soon after purchase. Trade in these species would then be banned.

Holthus says this would be premature, as the data are not available on which to make an informed choice. The first priority, he says, is to generate such data.

The current plan calls for certified animals to be offered to key markets in the United States, the largest importer of marine aquarium animals, beginning in about September 2001. ■

... as scientists raise alarm over coral reefs

Peter Pockley, Bali

Unusually for a specialist research conference, last week's ninth International Coral Reef Symposium in Bali, Indonesia, has pushed scientists to the forefront of the public debate on the effects of climate change on coral reefs at local, regional and international levels.

A consensus statement was triggered by an impassioned call from Yossi Loya, a veteran coral researcher from Israel's Tel Aviv University and winner of the Darwin Medal of the International Society for Reef Studies: "As a coral reef society we must add our voice to the growing international concern on the issue of global climate change, and call for an effective reduction in greenhouse gas emissions over the next decade."

After being given evidence for the dramatic decline of the world's reefs (see *Nature* 407, 932; 2000), the 1,430 delegates felt they could not wait the four

years to the next symposium, in Okinawa, Japan, before raising the alarm. Even those who only a year ago were sceptical that global warming is linked directly to mass bleaching events among corals have joined the chorus.

A scientific panel of symposium chairs and key speakers from four nations concluded: "Reefs face a bleak future as all climate models show that current increases in sea temperature will continue."

The panel agreed that there is "insufficient evidence that corals are able to acclimatize or adapt fast enough to these sort of changes," and nominated this as a priority area for research. But they added: "It seems perilous to use this as reason for little or moderated action."

But many delegates insisted that destructive fishing practices are both more influential and more immediately manageable than global warming. Concerned that

the issue might be forced to take a back seat to what they see as the 'sexy' issue of global warming, Mark Erdmann of the US Agency for International Development, among others, presented evidence that blast and cyanide fishing are "indisputably the biggest threats to coral reefs" off Indonesia (see opposite).

Erdmann says failure to apply strict regulations against these practices is costing Indonesia more than US\$200 million annually and threatens the world's greatest coral biodiversity.

In an unexpectedly frank speech, Indonesia's Vice-President Megawati Sukarnoputri admitted that only 6.5% of reefs in the Indonesian archipelago are "still in good condition", and a large part of the rest are "gravely damaged".

The case for urgent action will be taken directly to Washington and New York next month by Clive Wilkinson of the Australian Institute of Marine Science. ■

CERN physicists want more time to look for Higgs boson

Geneva The observation of a new 'event' that could be explained by the presence of the elusive Higgs boson has increased the probability that scientists at CERN, the European Laboratory for Particle Physics, really have seen the particle.

It has also led to demands that CERN's Large Electron-Positron collider (LEP), whose planned closure one month ago was postponed until this week to follow up earlier hints that Higgs had been spotted (see *Nature* **407**, 118; 2000), be kept running still longer — some suggest for up to a year.

Scientists working on the four experiments that are dedicated to searching for the subatomic particle will tomorrow (3 November) present the data they have gathered during the collider's one-month stay of execution. They are particularly excited about the new event because it was observed in an experiment where no events had previously been seen.

The researchers hope that the data will persuade CERN management to keep the collider operating. The choice is a difficult one, as an extension could lead to potentially expensive delays in the construction of the LEP's high-energy successor, the Large Hadron Collider. A final decision is expected next week.

Global warming happening faster than predicted

Washington Global warming could be happening more rapidly than previously estimated, leading to an average temperature increase of as much as 6 °C over the next century, according to the latest assessment by the Intergovernmental Panel on Climate Change (IPCC).

The report, which was leaked in advance of its expected completion in January next year, predicts that global warming will be greater than the IPCC's earlier assessments in 1990 and 1995. The panel had previously estimated the maximum likely temperature increase at around 3 °C.

The IPCC is also expected to be more assertive in its language, attributing the effect to greenhouse gases. Governments meet in two weeks time in The Hague to try to iron out the details of the Kyoto Protocol, the agreement that they reached three years ago to reduce greenhouse gas emissions.

Wings hold key to flexible boat mast

London It is not yet plain sailing, but a design of boat mast based on the bone structure of bat and bird wings has the finish line in sight



Flying high: Richard Dryden took inspiration from wings for his mast design.

after winning £45,000 (US\$65,300) from the UK's National Endowment for Science, Technology and the Arts.

The funding will help University of Plymouth biologist Richard Dryden to develop his transitional sailing rig, which changes shape according to wind conditions. Extended fully in light breezes, the jointed mast and sail becomes more swept back and flexed when the wind picks up.

Europe goes public over software patenting

Brussels The European Commission has launched a public consultation to help it handle the political and economic hot potato of whether computer software should be patentable.

The commission believes that Europe's lack of harmonized legislation on the issue may be a potential barrier to industrial growth, competitiveness and the development of the internal market.

The issue has prompted considerable debate in Europe. Supporters say software patents, such as those allowed in the United States, stimulate innovation. But critics claim they stifle competition. The consultation document is available on the commission's website until 15 December, and a decision is expected early next year.

► http://europa.eu.int/comm/internal_market/en/intprop/indprop/softpaten.htm

NASA plans smarter Mars landers

Washington As part of its revamped programme for exploring Mars, the American space agency NASA last week announced plans to develop 'smart' landing technology by 2007.

"Previously, spacecraft, like Mars Pathfinder, had no control over trajectory," says Firouz Naderi, head of the Mars Program Office at NASA's Jet Propulsion Laboratory in Pasadena, California. "We're looking at a class of lander that will recognize any deviations from the desired path and make corrections to get back on course."

The spacecraft should also be able to

ensure it reaches its destination in one piece. "The smart lander will be able to tell, using radar or some sort of camera, whether there are any ominous hazards directly below it," Naderi says, "and it will have the capability to move laterally as much as a football field length away, so it can land safely."

Salmonella pioneer wins Germany's top science prize

Munich Microbiologist Stanley Falkow of Stanford University this week won Germany's top science award, the Robert Koch prize, for his work on *Salmonella*, which has led to a new understanding of how pathogens spread. The annual prize is intended to promote the study of infectious diseases and other illnesses.

At the same ceremony, Swiss biologist Marco Baggiolini received the Robert Koch gold medal for his research efforts in identifying blood proteins that help the body fight off bacteria and viruses. Baggiolini, director of the University of Bern's Theodor Kocher Institute, is an acknowledged pioneer in the microbiology of the immune system.

US librarians worried by publishing merger

Washington US librarians have written to the Justice Department seeking the government's intervention to stop the merger between the European journal publisher Reed Elsevier and its US rival Harcourt General. The latter's range of publishing businesses include Academic Press, which will be merged with Elsevier Science.

According to Elsevier, the deal announced last week, under which it has paid \$4.5 billion for Harcourt General, will allow joint sales to grow "significantly faster" than the predicted 4–5% growth rate in demand for scientific information. The Association of Research Libraries is concerned that, by creating a concentration of power in the market for high-quality biomedical journals, attempts to achieve this goal will fuel a significant increase in subscription prices.

More money promised for Russia's research institutes

Moscow Yuri Osipov, president of the Russian Academy of Sciences, last week told the academy's praesidium that the government has promised the organization an additional 53 million rubles (US\$2 million) before the end of the year.

The money comes out of the 800 million rubles that the government has earmarked for optional expenditure, rather than committed expenditure such as salaries. The extra money will arrive at the peak of subscription renewal demands for scientific journals.



In focus: the Very Large Telescope can be configured to act as a 16-metre mirror.

Does size matter?

Large telescopes are starting to dominate astronomy, putting their smaller predecessors under pressure to close. But if astronomers can adapt their ways of working, says Alexander Hellemans, both big and small could thrive.

On remote mountain tops across the world, a new breed of optical and infrared telescopes stands sentry, watching the heavens. They are the colossi of modern astronomy. Twice the size of their predecessors, they are giants equipped with mirrors measuring up to ten metres across.

The first was the 10-metre Keck I telescope, sitting at the summit of Mauna Kea in Hawaii. It began science operations in May 1993 and was joined three years later by its partner, Keck II. Other behemoths are following close behind. "Until two years ago, we had only one or two large telescopes," says Shri Kulkarni, an astrophysicist at the California Institute of Technology (Caltech) in Pasadena. "A few years from now we will have almost ten."

This year, the eight-nation European Southern Observatory (ESO) finished building its Very Large Telescope on Mount Paranal in Chile. This consists of four 8-metre mirrors that can be operated independently or as a single telescope with the light-gathering capacity of a 16-metre mirror. On La Palma, one of the Canary Islands, Spain's 10.4-metre Gran Telescopio Canarias should begin routine operations by 2003. And on Mount Graham in Arizona, the Large Binocular Telescope, with two 8.4-metre mirrors, should be working by 2004.

But these giants do not come cheap. The Keck telescopes cost around US\$100 million each, and as astronomers strive after ever-larger mirrors, costs will continue to rise. A team at the ESO's headquarters in Garching, Germany, has hatched a plan for the Overwhelmingly Large telescope, which will have a mirror measuring 100 metres across. The project's leader, Roberto Gilmozzi, is confident that the team can keep the price tag for this telescope within the bounds of reason. But even so, it could not be built for much less than US\$1 billion.

Not surprisingly, budgetary constraints mean that some tough questions are now being asked. Even in the United States, where philanthropic institutions such as the W. M. Keck Foundation have bankrolled large telescope projects, some astronomers are questioning the wisdom of keeping the entire current roster of smaller telescopes in operation. And in Europe, where large and small telescopes compete directly for the same pots of taxpayers' money, the debate is particularly intense.

At its heart lie difficult questions about the relative scientific productivity of different telescopes. Astronomers cannot agree, for example, on whether it is better to operate one 10-metre telescope, or a series of 2-metre devices with the same light-collecting

area. But it is becoming clear that, for small telescopes to survive, they need to become more technologically sophisticated and must be deployed more effectively. The era of individual astronomers competing for time on these telescopes for their own pet projects could soon be over.

A problem resolved

At the most simplistic level, bigger is clearly better. Large mirrors gather more light, and so allow astronomers to observe fainter objects. They also have better resolution, giving sharper images. But there has been a gulf between theory and practice. Most ground-based optical and infrared telescopes have not achieved their maximum resolution because of the distortion effects caused by atmospheric turbulence. As a result, the Keck I telescope could not match the Hubble Space Telescope's resolution, despite the latter using a mirror just 2.4 metres across.

But adaptive optics could change all that. Developed by the US military¹ and by astronomers cooperating with industrial groups in France², adaptive optics detects the effects of atmospheric turbulence on a 'guide' star. It then uses this information to compensate for the distortion by deforming a flexible mirror onto which light collected from the

Torn between a rock and a high place

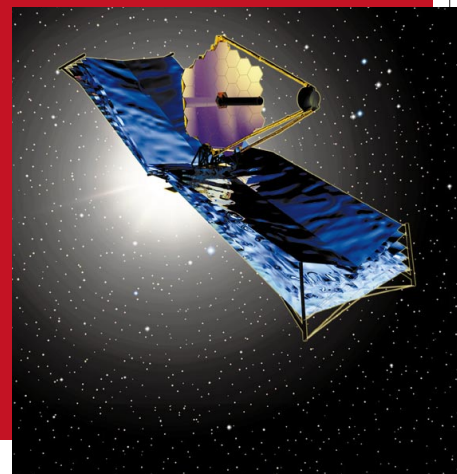
As astronomers mull over the merits of large versus smaller ground-based telescopes, another — equally contentious — debate is taking shape. Which will provide the best value for money: orbiting successors to the Hubble Space Telescope, or ground-based titans with mirrors of up to 100 metres in diameter? The answer may depend on whether adaptive optics, which eliminates the distortion caused by atmospheric turbulence (see main story), can be applied to the huge ground-based telescopes that are now on the drawing-board. “Adaptive optics is the single critical-path technology that has to be proven. And without high-quality wide-field adaptive optics, it is probable that, ultimately, there is no future for astronomy on the ground,” says Gerry Gilmore, an astronomer at the University of Cambridge.

Applying the technology to monsters such as the European Southern Observatory’s proposed 100-metre Overwhelmingly Large (OWL) telescope

poses a formidable challenge, not least because it will require supercomputers hundreds of times faster than current models. But if the technical obstacles can be overcome, OWL could provide images of individual stars within galaxies so far away that the light started on its journey to Earth when the Universe was a quarter of its current age. It could also observe planets orbiting around other stars in our Galaxy. Such performance would exceed by an order of magnitude that of the 8-metre Next Generation Space Telescope (NGST, right), which NASA, in collaboration with the European Space Agency and Canada, plans to launch around 2009, for about the same projected cost of US\$1 billion.

Indeed, the huge costs of getting large mirrors into space — not to mention the technical difficulties — mean that space telescopes will be unable to compete at wavelengths that can be studied from the ground, if the issue of adaptive

optics can be cracked. Space telescopes also have the disadvantage that their instruments cannot readily be updated. “You are always flying old instrumentation,” says Piero Benvenuti, who is coordinating European planning for the NGST.



NASA

telescope’s primary mirror is reflected. The technique’s power was demonstrated when the 3.6-metre Canada-France-Hawaii telescope on Mauna Kea produced images of the galaxy NGC 7469 that matched the resolution of those from Hubble³.

Last year, the technology was applied successfully at Keck II⁴, and the first adaptive optics images from Gemini North, an 8.1-metre telescope at Mauna Kea, were released last month. The technique is now central to the planning of all the main large-telescope projects.

Some astronomers believe that combining large mirrors with adaptive optics will

make many small telescopes obsolete. “In my opinion there are too many small observatories that now are too expensive,” says Lodewijk Woltjer of the Haute-Provence Observatory in France. “You can’t expect to keep doing what you used to do and at the same time start up new things.”

The French government seems to agree. Last year, it decided to stop core funding to telescopes smaller than two metres. Only strong protests by astronomers prevented the telescopes from being shut down. But in the absence of central government funding, observatories now have to seek alternative income. Stressing their value for training astronomy students, some telescopes are now being funded by individual universities or local authorities. The Pic du Midi Observatory high in the Pyrenees, meanwhile, has opened its facilities to ‘scientific tourists’. “We now have only two national telescopes in France, one at Pic du Midi, the other at Haute-Provence,” says Roland Bacon, director of the Lyon Observatory, who is seeking alternative funding for his observatory’s 1-metre telescope.

In Britain, smaller telescopes are also coming under financial scrutiny, as the UK Particle Physics and Astronomy Research Council negotiates terms for Britain to become a member of the ESO. An announcement on whether Britain will join is expected soon. And depending on the precise terms of any membership deal, Britain might have to stop funding some smaller telescopes, such as the 4-metre Anglo-Australian Telescope at Siding Spring in New South Wales and some members of the Isaac Newton Group of Telescopes on La Palma.

Meanwhile, the ESO is itself threatening to axe some of the smaller telescopes on its site at La Silla in Chile unless other organizations

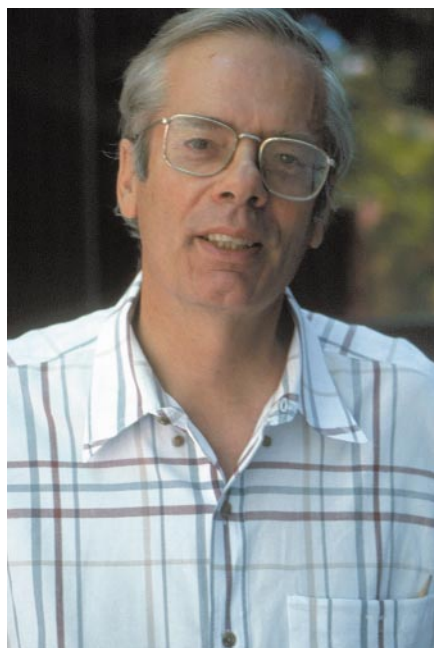
offer to take them over. “ESO is involved in facilities that are beyond the capabilities of a single country,” says Piero Benvenuti, an astronomer who works for the European Space Agency but is based at the ESO. “If a small telescope is only there for historical reasons, ESO should discontinue its use.”

The same questions surround many of the telescopes run by the US National Optical Astronomy Observatories — in fact, some of the organization’s smallest telescopes are already being transferred to consortia of universities. “There is a lot of hand-wringing about what to do with our national observatories,” says Roger Angel, whose work on mirror technology at the University of Arizona in Tucson is helping to drive the development of ever-larger telescopes.

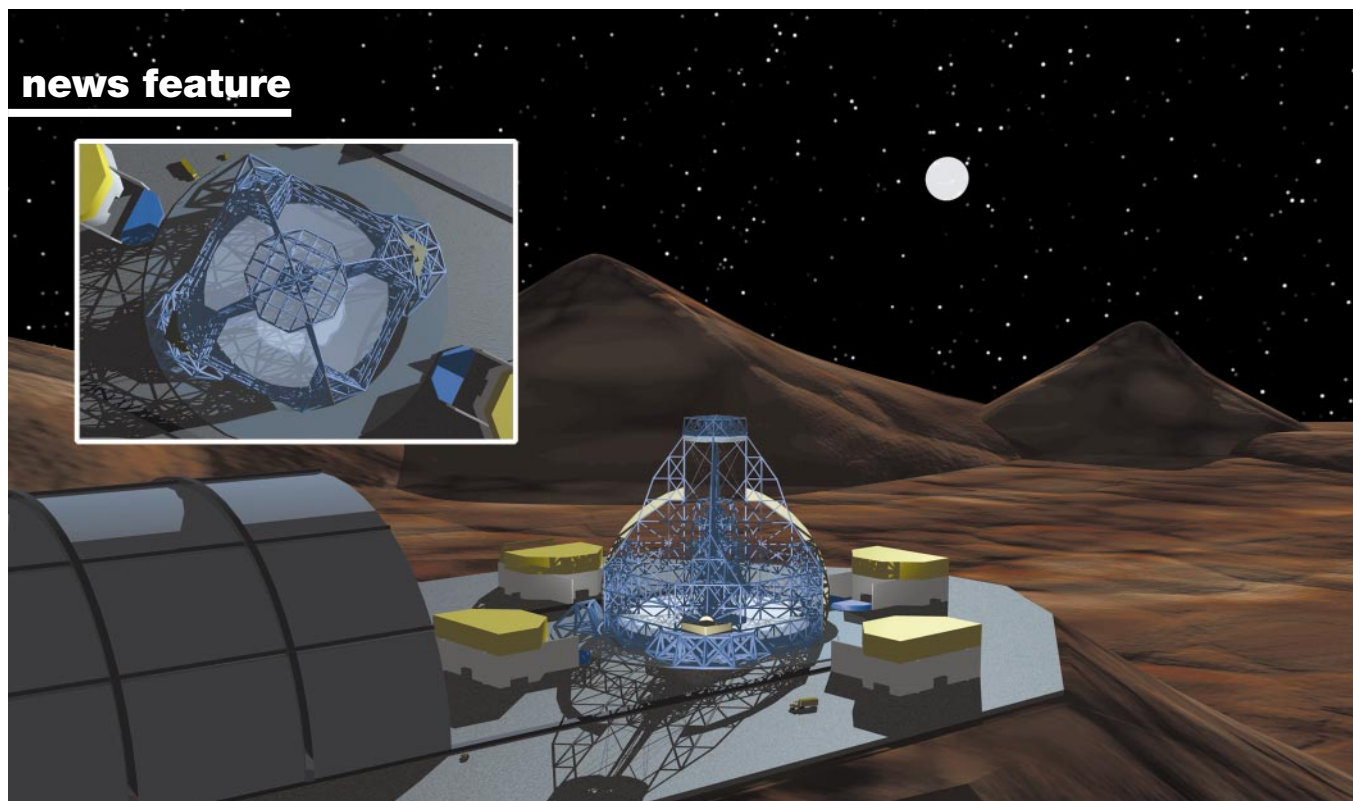
Small scopes take a stand

Some astronomers who use or work at the observatories now coming under threat feel that small telescopes are getting a raw deal, particularly as the costs of keeping them open are not huge. “A new telescope may cost US\$5 million a year to run, and the small telescope will cost US\$50,000,” says Helmut Abt of the Kitt Peak National Observatory in Arizona.

There are also fears that politics and egos are being allowed to dominate over a reasoned assessment of the scientific merits of the different telescope sizes. “Many of the people who are in senior positions have been using large telescopes,” says Richard Ellis, an astronomer at Caltech who is himself involved in a proposal to build a 30-metre device called the California Extremely Large Telescope. “In the political arena, the large telescopes get a lot of attention; they always get prominence in the newspapers. The



Mirror man: Roger Angel’s work is helping to drive the development of ever-larger telescopes.



► smaller telescopes deserve to be emphasized more than they have been," he adds.

Given these conflicts over priorities, some astronomers argue that the productivity of different telescopes should be subjected to rigorous quantitative analysis. Several researchers have attempted 'scientometric' studies over the years, counting publications or analysing patterns of citation in the astronomy literature.

Sizing up the issues

The most thorough study since the introduction of the Keck telescopes was done by Chris Benn, manager of the Anglo-Dutch 4.2-metre William Herschel Telescope on La Palma, and his colleague Sebastián Sánchez⁵. They analysed the 452 astronomy papers published in *Nature* between 1989 and 1998, and a sample of 1,000 papers making up the top 125 for citations in each of the years from 1991 to 1998. They reasoned that these papers represented the cream of the scientific output of the astronomy community, and hoped that analysing the contributions made by different telescopes would indicate the relative scientific productivity of different size classes.

A similar picture emerged both for the citation data and counts of *Nature* papers, with the Keck telescopes making an immediate impact from the mid-1990s. For papers published in 1995–98, for instance, the citation impact of Keck I was eight times greater than that of a typical 4-metre telescope, and the mean citation impact of telescopes in the 2-metre class was four times lower still. These ratios roughly reflect the differences in mirror collecting area between the telescopes. So although larger telescopes were more productive, they did not seem to have a disproportionate advantage.



Little and large: is there scope for all sizes of observatory, or will the planned Overwhelmingly Large telescope (top) leave smaller set-ups such as that at Pic du Midi out in the cold?

When Benn and Sánchez compared different size classes of telescope according to the total mirror surface area available within each class, the productivity of 2-metre telescopes held up well, even compared with the two Keck telescopes. "The strong showing by 1-metre and 2-metre telescopes in the 1990s augurs well for the continued scientific impact of 4-metre telescopes in the era of 8-metre telescopes," Benn and Sánchez's paper concludes. Whether this conclusion will hold up once adaptive optics is in widespread use remains to be seen. But for now, Benn argues that "it is very important to maintain a variety of telescopes of different sizes and types".

Other researchers point out that certain subdisciplines within astronomy are particularly suited to smaller telescopes. In fact, one of the most famous astronomical discoveries of recent years, the detection of the first extrasolar planet around a Sun-like star, was made by Michel Mayor and Didier Queloz of the Geneva Observatory using the 1.9-metre telescope at Haute-Provence⁶. Most of the subsequent discoveries in this area have been made using telescopes with mirrors less than three metres in diameter.

Detecting the 'wobble' of a star caused by an orbiting planet does not require a large telescope. But it does need up to 50 nights of observations, which are impossible to obtain

Europeans ponder the scope for improvement

For any funding agency, the issue of deciding which small telescopes to keep and which to shut down is fraught with difficulties. But the US National Optical Astronomy Observatories, based in Tucson, Arizona, at least has the advantage of controlling a suite of observatories serving astronomers across an entire continent, making it possible to consider the provision of telescopes within an integrated system.

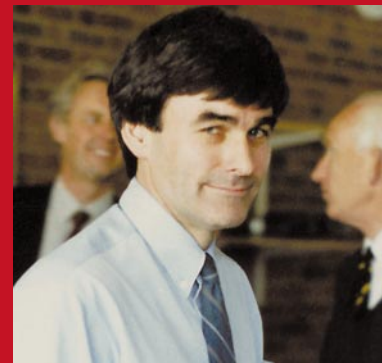
In Europe, where individual nations run a handful of telescopes each, it is much harder to get an overview of how best to serve the continent's astronomers. "If each country keeps operating its own telescopes independently, you will end up with a large number of telescopes doing substandard work," says René Rutten, director of the Isaac Newton Group of Telescopes on La Palma in the Canary Islands. He argues that European nations should close many of their existing small telescopes and run the remainder jointly, concentrating on

developing the best site for optical and infrared astronomy in Europe — La Palma.

For now, such radical moves are not under serious discussion. But in March this year, an organization called OPTICON (Optical and Infrared Coordination Network for Astronomy) was established in an attempt to improve the organization of astronomy at the European level. "We look at new ways of making sure that the small telescopes do what they are good at in a cost-effective way," says Gerry Gilmore, an astronomer at the University of Cambridge, who chairs the group.

Funded by the European Commission to the tune of 1 million euros (US\$836,000) over three years, OPTICON has neither the budget, nor the political influence, to enforce better coordination. But Gilmore is optimistic that national agencies will see the merits of eliminating "needless duplication". That streamlining process will start in

earnest at OPTICON's first major meeting, to be held in Lyon in December. In the long run, says Gilmore, "it's highly likely that we will have a smaller number of more efficient telescopes".



Gilmore: believes improved coordination will give small telescopes a better focus.

on one of the giants. "Work with a small telescope is much cheaper, and people can get much more observing time," says Abt.

Any project that needs to make repeated observations of an astronomical object depends on small telescopes for this very reason. Brian Marsden of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, who leads a team that acts as a clearing-house for data about comets, asteroids and other objects in the Solar System, adds that small telescopes continue to dominate his field. "There are very few observations of anything that comes our way made with a telescope larger than four metres," he says.

In reflective mood

Nevertheless, limited budgets mean that the least productive of the small telescopes currently in operation may face the chop — or at least will get taken out of the front line of research and be used for training. To avoid that fate, the operators of small telescopes could do worse than to imitate the current star performers within the medium and smaller size classes.

In Benn and Sánchez's citation analyses, the Canada-France-Hawaii telescope was the most productive telescope in the 4-metre class. It combines two attributes that are likely to emerge as the keys to survival. These are a strong focus on developing technology such as instrumentation, and a policy of concentrating on a particular scientific problem, rather than the usual practice of allowing astronomers working on a plethora of projects to compete for slots of observing time.

"It doesn't pay to let everybody and his dog use your telescope," says Virginia Trimble of the University of California at Irvine, who noted the high productivity of the

Canada-France-Hawaii telescope in an earlier scientometric study⁷. This telescope not only helped to pioneer adaptive optics, but its operators also took the decision to focus primarily on assigning redshifts to large numbers of objects, so determining their distances from the Earth.

Among the telescopes with mirrors smaller than two metres, Benn and Sánchez identified the 1.3-metre telescope at the Mount Stromlo Observatory near Canberra in Australia as the strongest performer. This has also been used for a specific purpose — the Massive Compact Halo Object project, which is searching the halo of our Galaxy for dark matter in the form of 'failed' stars called brown dwarfs⁸.

Caltech's Kulkarni is convinced that these examples point the way to the future for small telescopes. "The strategy for the coming decade is to take these old telescopes and do very focused experiments with them — a creative new style of doing astronomy," he says. But Kulkarni warns that established ways of working, which do not readily lend themselves to such focused collaborations, are entrenched among researchers who are used to letting other people run the telescopes while they just compete for observing slots. "Most astronomers are still thinking in an old-fashioned way — 'your time, and my time,'" he says.

Adapt and survive

If small telescopes are to be adapted to specialize on particular scientific problems, they need to be fitted with specific instrumentation to optimize them for the task in hand — rather than the 'one size fits all' instrumentation typically fitted at the moment. And that, says Carl Akerlof of the University of Michigan, may require a shift

in emphasis on the part of funding agencies, and in attitudes among astronomers. "Scientists who get involved in instrumentation are regarded as lower in the pecking order," says Akerlof, who believes that this view delayed the development of adaptive optics by "many years".

Akerlof heads a project which suggests that new small telescopes will still be around even when the current generation is too old to be worth maintaining. His Robotic Optical Transient Search Experiment (ROTSE), based at the Los Alamos National Laboratory in New Mexico, last year captured the first-ever optical images of a gamma-ray burst as it occurred⁹. In terms of optics, ROTSE consists of nothing more sophisticated than an array of four telephoto lenses. The key to its success was again a combination of a narrow scientific focus and the latest technology — in this case, automated operation that allowed ROTSE to home in almost instantaneously on the burst, when it was triggered by data received from NASA's orbiting Compton Gamma-Ray Observatory.

Astronomers who fear that mammoth telescopes are about to consume their field's financial resources to the exclusion of other observatories have good justification for arguing that size is not everything. But if their voices are to be heard, they will need to heed the other half of that saying: "It's what you do with it that counts."

Alexander Hellemans is a science writer in Naples.

1. Hardy, J. W. *Proc. SPIE* **332**, 252–259 (1982).
2. Roddier, F., Nortcott, M. & Graves, L. N. *Publ. Astron. Soc. Pacif.* **103**, 131–149 (1991).
3. Rigaut, F. *Publ. Astron. Soc. Pacif.* **110**, 152–164 (1998).
4. Wizinowich, P. *Publ. Astron. Soc. Pacif.* **112**, 315–319 (2000).
5. Benn, C. R. & Sánchez, S. F. *Publ. Astron. Soc. Pacif.* (in the press).
6. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
7. Trimble, V. *Scientometrics* **36**, 237–246 (1996).
8. Alcock, C. et al. *Astrophys. J.* **486**, 697–726 (1997).
9. Akerlof, C. et al. *Nature* **398**, 400–402 (1999).

Genes should not be patentable, but if they are, we all have to do it

Sir—Public debate about advances in molecular biology and the billion-dollar revenues that may accrue from new therapies has taken on particular significance in Brazil. This is a result of the enormous success of the consortium which reported the genetic sequencing of *Xylella fastidiosa*¹, the bacterium causing the 'amarelinho' disease that destroys more than 30% of São Paulo's orange groves.

The same consortium is making great sequencing progress in three other areas: human cancer genes endemic to Brazil; genes related to sugar-cane metabolism; and the genome of the *Xanthomonas citri* bacterium which causes citric cancer, widespread in Brazil's orange groves.

Brazilian discoveries need to be protected, but the country's legislation does not permit the patenting of live beings. Brazilians are inexperienced in dealing with international patents. Of the 100,000 or so patents granted by the United States patenting office each year, a mere few dozen go to Brazilians.

Every original invention that is useful and has a commercial potential can be patented somewhere in the world. Human imagination knows no bounds: take, for example, US patent 5443036, a device for encouraging a cat to exercise by chasing a light spot². However, the US statutes define four types of inventions for purposes of registration: new processes (or methods); machines (or devices); manufactured articles; and new compositions of matter.

Processes or methods are inventions that describe how to do something. A typical claim could be "a method to make vegetable soup", which comprises all the necessary steps for preparation. Devices are machines that do something, such as "brush teeth automatically", followed by a description of the parts that make them up and how they interrelate. An example in the third category might be an "optical fibre", with a detailed description of its structure and composition.

In the fourth category, although matter existing spontaneously in nature is not patentable, new compounds and chemical compositions can be. Thus, a synthesized bioactive glass for the substitution of bones and teeth, containing oxygen, silicon, sodium, calcium and phosphorus, has been patented³. But its constituent elements, being natural, cannot. (One can imagine the possible consequences had each element of the periodic table been patented following discovery.)

A patent requires human intervention

in the design, construction and synthesis or manufacture of the product. Another requirement is descriptive sufficiency: enough information for the invention to be reproduced. Hence there is a clear difference between invention and discovery. The former results in a new composition, product, device or process; the latter from the unveiling of universal laws or from the structure or composition of extant natural matter.

Powerful lobbies have encouraged the filing of patents on live beings. North American and European agencies — with the exception of the French — have granted several thousand patents for genes and genomes, despite the promises of world leaders. The most common argument is that patents will be granted only after the gene's functionality has been clearly established. But surely this is a question of discovery, not invention?

From an ethical standpoint, genes should not be patented (although medical inventions based on such a discovery can and should be). Yet in industrialized countries researchers are filing thousands of gene patents: Celera Genomics alludes to 6,500. So how should Brazil, or any country, protect the public resources that make these discoveries possible? Until logic and common sense prevail, we should patent our inventions everywhere. However, one could deposit discoveries that might have immediate commercial interest in electronic databases and charge for privileged access. This would not prevent other investigators continuing their research, even if they could not afford privileged access to the latest discoveries.

Edgar Dutra Zanotto

7-09-00 Federal University of São Carlos, 13.565-905, São Carlos, São Paulo, Brazil

1. *X. fastidiosa* Consortium *Nature* **406**, 151–157 (2000).
2. Amiss, K. T. & Abbott, M. H. Method of exercising a cat. US Patent 5443036 (1995).
3. Hench, L. *et al.* Bioactive ceramics and method of preparing bioactive ceramics. US Patent 5981412 (1997).

Results may not fit well with current theories ...

Sir—There is a narrow inevitability about Partridge and Barton's response¹ to True and Lindquist's evolutionary interpretation of their remarkable finding: that the presence of the yeast prion $[PSI^+]$ releases cryptic phenotypic variation which allows cells to thrive in fluctuating environments and which may facilitate the establishment of new traits².

Partridge and Barton hover between dismissal and marginalization — the last resting place of many new discoveries of genome and developmental dynamics that do not fit easily with the simplistic notion

of so-called 'modern' Darwinism that DNA mutation and recombination are the sole heritable changes of state of concern to selection.

What is at stake is whether heritable prion-induced variability in protein size is just a "side effect of disrupted gene expression" or an evolved system for the release of phenotypic variation. Have biological systems evolved evolvability?

The yeast prion case is but the latest example in a long list of ways by which organisms increase variation during the intricate processes of transforming genotypic information into phenotypes. There is no need to remind *Nature's* molecular genetics readers of the total phenomenology of differential promoter utilization, DNA modification, differential DNA and RNA splicing, RNA editing, and post-translational modification. To these, we can now add Lindquist's earlier report of stress-induced release of variation via heat-shock protein chaperones³ and the $[PSI^+]$ story.

All such systems — such as the DNA-based systems of recombinational variation released by sex, the proof-reading systems involved with DNA replication, and genetic redundancy resulting from genomic turnover — have surely been influenced willy-nilly by selection to produce an exploitable balance between the eternal contrasting needs of stasis and change⁴.

The complex systems of proteins and RNA that are part of such multiple checks and balances are not just side effects of something else supposedly more fundamental and selectable — they are of the essence of responsive developmental processes. Either all of these systems facilitate evolution or they all don't. Given our current ignorance, we can't pick and choose.

Gabby Dover

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

1. Partridge, L. & Barton, N. H. *Nature* **407**, 457–458 (2000).
2. True, H. L. & Lindquist, S. L. *Nature* **407**, 477–483 (2000).
3. Rutherford, S. L. & Lindquist, S. L. *Nature* **396**, 336–342 (1998).
4. Dover, G. A. *Dear Mr Darwin: Letters on the Evolution of Life and Human Nature* (Weidenfeld and Nicolson, London, 2000).

... but yeast prion offers clues about evolution

Sir—Partridge and Barton in their News and Views article¹ on our paper² make many interesting points, but they also misrepresent our hypothesis. Our main hypothesis is not that the yeast prion $[PSI^+]$ is maintained by natural selection because it aids evolution. This was the last of the three

suggestions in our paper, and the one we least favoured. Partridge and Barton propose, as if it were an alternative view, that the increased variation [PSI^+] produces is “a side effect of disrupted gene expression”. But this is obvious, and we never suggested otherwise.

The main point is that, no matter how it arose or is maintained, the prion has strong and remarkably varied effects on growth. This seems likely to permit survival in fluctuating environments and provides a plausible route to the evolution of new traits.

In our paper, we suggested that a capacity to accumulate silent variation and release it in a combinatorial fashion might facilitate evolutionary change. But this was not the motivation for our work, as suggested by Partridge and Barton. Rather, we asked if this very unusual form of inheritance might provide a selective advantage. It was the extraordinary diversity of the traits we observed that suggested broader implications. We never claimed these traits must be due to “large variations, involving several random changes”. Some might be, and this has important implications. But others might be due to something as simple as the addition of a few amino acids to a single kinase that increases its stability or prevents association with a repressor.

Finally, we did not suggest that “there is no path by which natural selection could construct [complex adaptations].” We are simply not convinced that stepwise, individually selected non-deleterious pathways are sufficient to explain all evolution.

When we discovered in earlier work that Hsp90 has a massive capacity to buffer morphogenetic variation and release that variation in response to environmental stress, we hypothesized that this protein might have a previously unappreciated impact on evolution, as a natural consequence of its special role in protein folding. We did not, however, as has been suggested³, claim that these properties evolved to this end.

It is disappointing that in many US schools the doctrine of creationism is given equal weight with the theory of evolution (see, for example, ref. 4). Plausible explanations for the sometimes puzzlingly rapid pace of evolution may help to counter arguments that evolution cannot have done what it is held to have done.

Susan Lindquist

Department of Molecular Genetics and Cell Biology, Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA

1. Partridge, L. & Barton, N. H. *Nature* **407**, 457–458 (2000).

2. True, H. L. & Lindquist, S. L. *Nature* **407**, 478–483 (2000).

3. Dickinson, W. J. & Seger, J. *Nature* **396**, 336–342 (1998).

4. Lerner, S. L. *Nature* **407**, 287–290 (2000).

In defence of Spanish R&D spending

Sir—Your editorial on the 2001 Spanish budget for science and technology (*Nature* **407**, 659; 2000) contains some claims that are far from the real facts. This article, written in sensationalist language, has created much confusion in the Spanish press and in the scientific community, where a claim from *Nature*—thanks to its well-deserved scientific reputation—is taken as the truth.

The editorial mentions the Spanish government’s “secrecy” and says I was “reluctantly forced to admit that more than 50% of our 2001 R&D budget will be devoted to military expense”. First, there is no secrecy or reluctance. The budget figures—including military research expenditures—are in the public domain, and I have explained them in the Congress, in the Senate and in press conferences.

Second, I never said that over half of our R&D budget is spent on defence, as this is simply not true.

It is misleading to mix expenditures on grants in with government loans—money which will one day be returned. Spain’s 2001 budget for research and technological programmes will be managed mostly (85.5%) by the new Spanish Ministry of Science and Technology (MST)—only 9.3% will be controlled by the Ministry of Defence. In the MST, the projected research budget, excluding loans, is 1,588 million euros, of this exactly zero will be used in defence programmes. Defence expenditures appear only under our government loans—at zero interest—programmes. Here the total is 2,563 million euros, of which 1,447 million euros can be associated with technological investments in defence. These numbers are large in relation to funds allocated to grants, but the implied subsidy is only in the interest-rate differential (and loan maturation) and not the total amount. Defence expenses within our ministry budget will grow at 5.7%, whereas basic research programmes will grow by 10.1%.

One could simply spend less on R&D defence programmes to spend more elsewhere. This is how things were done in the past. However, the “elsewhere” was geographical, as defence technology was imported from abroad. This situation has changed radically over the past few years. To a large extent, thanks to the support of R&D defence programmes, Spain now has an innovative, competitive and fast-growing aeronautics industry that has created new products and jobs.

There are, however, other aspects of the budget that are new. This is the first year since Spain regained democracy that there

is no budget deficit. Nevertheless, R&D expenditures are growing substantially above the average government expenditures of 4.5%. Priority is given to increasing the number of research positions and long-term contracts (more than 800 in 2001, to achieve the 2,000 goal by 2003). Programmes devoted to human resources are planned to grow at 18%, and a further plan will be launched to attract Spanish researchers back from abroad.

One would have expected that *Nature*, having been concerned about these issues in the past, would have emphasized these priorities and innovations.

Ramón Marimon, Secretary of State for Scientific and Technological Policy, Ministry of Science and Technology, Paseo de la Castellana 160, E-28071, Madrid, Spain

Don’t ignore the risk of vaccine contamination

Sir—Your News and Opinion articles^{1,2} about alleged contamination of vaccines should serve as a warning against over-optimism.

These articles highlight the failure to show any evidence for contamination of Wistar Institute polio vaccine stocks by human and simian immunodeficiency viruses (HIV/SIV), and you appeal for a truce. But—although Edward Hooper is quoted as saying that “vaccine samples released did not include any from batches prepared for use in Africa”—lymphocytes have been detected in other polio vaccines³. Half of the vervet monkeys in Southern Africa are SIV positive; these animals were used for preparing early polio vaccines.

Considering the many millions of vaccine doses prepared in primary vervet monkey kidney cultures over a 30-year period, it is inconceivable that some SIV did not contaminate many cultures. By the same yardstick, simian virus 40 (SV40) contaminated millions of doses of poliovirus vaccine until the animals were screened for this tumour virus.

Edward Hooper and others surely do not intend to undermine the polio vaccine efforts. What is needed is a new awareness of the need for caution—remembering the example of BSE—in view of the current impetus towards xenotransplantation and the accompanying danger of contamination. Our aim should be to improve our vaccines, not to undermine public confidence in them.

G. Lecatsas

Department of Virology, Medical University of Southern Africa, Medunsa 0204, South Africa

1. *Nature* **407**, 115 (2000).

2. Dickson, D. *Nature* **407**, 117 (2000).

3. Ohta, Y. et al. *AIDS* **3**, 183–184 (1989).

Is Russian science recovering?

A Nobel Prize raises morale, but research needs money and democracy.

Irina Dezhina and Loren Graham

After a precipitous decline following the Soviet Union's collapse in 1991, Russian science may be on the slow road to recovery — or at least stabilization. The new president, Vladimir Putin, has placed strong emphasis on science and technology, although many Russian fundamental researchers are worried about his approach. They fear that he is reducing science to practical military and industrial interests, that he is ignoring environmental concerns, and that he may be reverting to centralized planning of scientific and technological developments, rather than encouraging the healthier approach of growth from below. It is still too early to tell whether Putin will match his positive rhetoric with actual budget increases.

Catastrophic drop

Between 1990 and 1998, the budget for Russian science fell to about a quarter of its 1989 value and the number of active researchers fell from almost a million to about 400,000. Thousands of scientists emigrated, including some of the most prominent; many more were part of an 'internal brain drain' as they transferred to more lucrative activities in the private sector where their scientific training often was not used. Within a few years, the world's largest scientific and technological enterprise had withered catastrophically^{1,2}.

Putin was elected president in May this year, and the same month he gave a speech to the Academy of Sciences, the traditional home of fundamental research, in which he stated: "Science is the most important source of economic development and the safeguarding of the national security of the country." His priorities in science policy, he said, were to support national defence, to stimulate innovation, to attract talented young people into scientific careers, and to supply fundamental science with new equipment and better telecommunications.

Putin's first step was the creation of the Ministry of Industry, Science and Technology, making it one of the largest federal structures. It has more than 1,400 enterprises under its direction, mostly in the military-industrial complex³. At the same time, he liquidated the older Ministry of Science and Technology and several other federal organizations, including the State Committee on the Environment. Some people, including the St Petersburg Association of Scientists, worried that science was being converted into a 'service branch' for military issues⁴, and that fundamental and environmental



Putin (centre) claims to be backing Russian science, but can he deliver on his promises?

science were being downgraded.

The new science minister, Alexander Dondukov (previously general director of the Yakovlev aviation company), immediately announced his intention to prevent further erosion of Russia's scientific potential and to emphasize information technologies, including the development of a supercomputer. He also presented the rather murky goal of "codifying" Russia's scientific knowledge and experience by creating a register of all currently existing valuable technologies, a project that has its critics⁵.

Research pay is on the rise

The general economy in Russia is now growing after years of decline, with industrial production increasing by almost 10% in the past year, particularly because of revenue from oil

exports. As a result, for the first time in almost a decade, the average federal salary of a researcher now exceeds that of the average employee in the rest of the Russian economy, reaching an 18% advantage in July 2000. In 1999, average pay for a researcher was \$68.70 a month; in the first months of 2000 it was \$80.40 — but this is still below the salary of the average industrial worker.

Some scientific researchers have increased their income further by relying on non-federal sources. Regional governments are funding about 13% of the country's scientific research. European and US foundations are increasing the size (but reducing the number) of their grants to Russian research. These foundations are becoming more willing to support institutions, rather than just individuals, by meeting overhead costs. An example is the support given to the Engelhardt Institute of Molecular Biology in Moscow by the Howard Hughes Medical Institute.

Support for the young

Even though a few 'oligarchs' — the businessmen who made fortunes out of privatization — are now very wealthy, philanthropy has not emerged as a significant private source for scientific research. One reason is that Russian foundations are taxed but for-

The position of science in Russia is more hopeful now than at any time since the collapse of the Soviet Union.

Foreign foundations are not. Nonetheless, some private philanthropy is beginning to appear, especially in education. For example, the Russian Credit Bank, the Interros Foundation (supported by the oligarch Vladimir Potanin) and the oil company Yukos are all making grants to university departments⁶.

The support of young researchers is generally recognized as the most crucial challenge facing Russian science. In the past five years the total number of researchers has decreased by a fifth, but among those under 30 years of age this loss rises to a third and between 30 and 40 years of age to 40%. By the end of this year, there will be six Russian scientists at or above retirement age for every five who are less than 40 years old⁷. Recognizing this threat, several government agencies and scientific organizations, including the Ministry of Education, the Russian Academy of Sciences and the Russian Foundation for Fundamental Research (a post-Soviet government organization modelled on the US National Science Foundation), have announced new programmes, prizes, competitive grants and fellowships, special living quarters, and exemption from military conscription. The Russian Academy of Sciences is assigning 130 million rubles (US\$4.7 million) per year to help young scientists.

One of the major flaws of the Soviet system of higher education and science was the rigid separation of research and teaching. The Academy of Sciences was primarily responsible for fundamental research, but did little teaching; the universities were mainly responsible for teaching, but did relatively little research. In 1997, the Russian government began an integration programme by inviting Academy researchers to lecture in universities and by allowing students to work in Academy institutes, which often have superior equipment. The Academy, after earlier misgivings, is now supporting this project, in the hope that it can identify the best students and invite them to work in its institutes, providing the fresh blood that it so badly needs.

But the government has now decided to include any scientific-technical organization in the competition for this programme, which could cause problems if ambitions expand but funds remain relatively modest (according to data in late 1999, \$11.4 million was designated for this programme). Foreign foundations, particularly the John D. and Catherine T. MacArthur Foundation and the Carnegie Corporation in the United States, have devoted large resources to promoting the integration of Russian research and teaching, for example in their Basic Research and Higher Education Programme in which eight Russian universities are slated to receive about \$1 million each.

Meanwhile, the Russian government is planning to designate 25 or 30 of the best existing universities as 'federal universities',

Whether Putin's strategy can fit well with true creativity in science and technology remains to be seen.

but no one knows yet just how much additional money they will receive. If sufficient funds can be found, and if the focus can be maintained, this integration programme may become one of the success stories of Russian science and technology.

The government is expanding on plans developed during Boris Yeltsin's administration to promote the creation of centres of innovation called, confusingly, "technoparks", "technological incubators", "science cities", and "federal centres of science and technology". It is also creating financial institutions to promote innovation, such as the Russian Foundation for Technological Development, the Federal Fund for Assistance to Small Innovative Enterprises, and the Venture Innovation Fund⁸.

In Soviet times there were about 60 science cities, or *naukogrady*—special defence-related metropolitan areas. After the collapse of the Soviet Union many of these cities experienced severe economic difficulties and high unemployment. Now Putin intends to give them legal status and privileged tax regulations, in the hope that they will develop along the lines of Silicon Valley in California and Route 128 near Boston. So far, only the city of Obninsk in the Kaluga region has achieved this new status, but many others aspire to the 25-year period of financial incentives and subsidies. The exact amounts of money to be devoted to these science cities is unknown; it appears that the major input will be tax exemptions. Perhaps this attempt to create science cities will succeed, but the California and Boston areas of innovation were created by entrepreneurs and companies choosing their locations for themselves, not by government decree.

Bright spots

Although Russian science continues to suffer from severe financial deprivation, some bright spots have appeared. The award of a Nobel Prize in physics this year to Zhores Alferov (admittedly for work done during the Soviet period) was a psychological boost, and Alferov appealed to the government to give more money for science in a speech to parliament the day after he received the prize⁹. He also noted that highly talented young people are today still entering fields in which Russia has traditionally been strong, such as physics and mathematics. Important work is being done in chemistry at Moscow

University, where the department of chemistry has received numerous foreign grants, and in metallurgy at the Moscow Institute of Steel and Alloys, which also has foreign cooperative programmes. At the Engelhardt Institute of Molecular Biology, significant research is being done on biochips and on a possible anti-AIDS drug, phosphazide, now being tested at McGill University in Montreal¹⁰. Space research, although greatly weakened by financial cutbacks, continues, illustrated by current cooperation at the international space station. Recognizing the high quality of Russian space and aviation engineers, Boeing operates its only foreign research facility in Moscow. These are just a few examples showing that, although science in Russia is suffering, it is far from dead.

In fact, the position of science and technology in Russia is more hopeful than at any time since the collapse of the Soviet Union. As the economy improves, so does funding of research and education. Some new sources of funding are appearing, along with a new culture emphasizing the commercialization of technology. Of course, the financial increases (so far, non-budgetary) for science are still small and economic expansion is fragile, depending on international factors that could change, such as the price of oil and the exchange rate of the ruble.

Furthermore, the Putin government seems to be reverting to central control, and military and industrial priorities disturbingly reminiscent of Soviet practices. Putin's concern about the state of Russian science and technology and his wish to improve it are to be welcomed, but whether his strategy can fit well with democratic government, a free economy, international cooperation and true creativity in science and technology remains to be seen.

Irina Dezhina is at the Institute for the Economy in Transition, *Gazetny per. 5, Moscow 103918, Russia.*

e-mail: dezhina@crdf.ru

Loren Graham is at Room E51-163, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mail: lrg@mit.edu

- Dezhina, I. *Rossiiskaya Ekonomika V 1999 Godu: Tendentsii I Perspektivy* 21, 615–630 (Institute for the Economy in Transition, Moscow, 2000).
- Graham, L. *What Have We Learned about Science and Technology from the Russian Experience?* (Stanford University Press, 1998).
- Nezavisimaya Gazeta* 4 (9 June 2000).
- Libenson, M. N. & Borkin, L. Ia. "Statement of the Coordination Council of St Petersburg Association of Scientists" (in Russian), unpublished petition, June 2000.
- Antokhin, M. "Alexander Dondukov promises a Russian supercomputer" <http://SMI.RU> (2 August 2000).
- Nezavisimaya Gazeta* 13 (10 March 2000).
- Calculated from statistics in *Nauka V Rossii V 1996 Godu* 18 (State Committee on Statistics of the Russian Federation, Moscow, 1996); and *Nauka V Rossii V Tsifrakh — 1999* 37 (CSRS, Moscow, 1999).
- Dezhina, I. & Tsyganov, S. *Vestnik Rossiiskoi Akademii Nauk* 4, 291–294 (2000).
- Wines, M. "Nobel winner laments poverty of Russian science" *The New York Times* A3 (12 October 2000).
- "What Russians have contributed" *Pittsburgh Post-Gazette* (31 May 1999).

A powerful leap from chaos

The natural world, financial crises, wars, history — all modelled in a pile of sand.

Ubiquity: Why the World Is Simpler Than We Think

by Mark Buchanan

Weidenfeld & Nicolson: 2000. 230 pp. £20

Niall Ferguson

A few years ago, historians belatedly caught up with the fact that science had moved on somewhat since Newton and Laplace. Years had been spent down culs-de-sac named Hegel, Marx, Toynbee and Braudel, in the vain pursuit of 'laws of history' comparable with the laws of motion. Whereas some historians had reversed out of these dead ends onto the ring road of literary criticism, abandoning the notion that history might have anything to learn from the natural sciences, a few ventured down the winding lane of chaos theory.

The key insights of chaos — sensitivity to initial conditions, stochastic behaviour in deterministic systems — had an obvious relevance to historical study, not least because one of the best-known chaotic systems, the world's weather, dominated the lives of most people throughout the long pre-industrial era.

As a historian, what I liked about chaos theory was the way it reconciled the manifest unpredictability of history — its *apparent* chaos — with the reality of determinism, in the sense of meaningful causal relationships. History was not just an infinite number of contingencies. Rather, the multiple equations at work in history were nonlinear. So teleological philosophies of history such as that of Marx could be consigned to the rubbish bin without the need to retreat to the extreme position of the idealist philosophers of history (and of the post-modernists) — that historians are engaged in nothing more than the construction of narratives whose meaning is as dependent on literary convention as on any empirical verifiability.

History was authentically chaotic: just as the famous Amazonian butterfly's wings could indirectly unleash a storm over England's Home Counties, so a single political decision could make the difference between war and peace. And precisely for that reason it was a worthwhile exercise to contemplate what might have happened if the decision had gone the other way: if the butterfly had not flapped its wings. 'Counterfactual' history — the asking of 'what if' questions and the contemplation of parallel worlds — is a lot easier to justify in a chaotic universe than in a universe of rigidly deterministic laws.

Now comes Mark Buchanan with solace for those discomfited by historical debates conducted largely in the subjunctive. "The



world," he tells us, is "simpler than we think." By this he means not only the natural world of earthquakes and biological evolution, but also the human world of financial crashes, scientific breakthroughs, international wars and political revolutions.

I must confess to feeling ambivalent about Buchanan's project. Part of me was not unhappy with the idea of 'chaostory', which created the opportunity to turn history into a series of subversive counter-narratives of what (to invert Ranke's dictum) did *not* 'actually happen'. That part was rather alarmed at the possibility of some new simplification imported from the natural sciences. But the other part of me — the part that is attracted to the empirical rigour of the social sciences, and which has spent the past two years poring over the fluctuations of the financial markets in the hope of discerning something other than a 'random walk' — was rather excited at the prospect of a return to linearity.

Buchanan is a science journalist with a PhD in theoretical physics. He does a superb job in this book of explaining a number of important related breakthroughs in the natural sciences, ranging from the incidence of earthquakes to mass extinctions. This is popular science writing at its best, translating recent research in physics and biology into language a layman — even a historian — can grasp.

The key concept — the 'ubiquitous' idea promised by his title — is the 'power law'. Nine out of 15 chapters are concerned with demonstrating the existence of this law in the natural world or, more commonly, in the simplified versions of that world constructed by ingenious physicists with the help of their computers.

Buchanan's starting point is a game

played with a pile of sand (or grains of rice). Drop individual grains, one after another, onto the pile and watch what happens — sometimes not much; but every now and then the pile collapses completely.

Statistical analysis reveals that there is no 'typical' or average sand-slide. Plotted on a graph, the distribution of slides is not the familiar bell curve. Rather, if you plot the size of slides against the frequency of their occurrence, you get a straight line. To use the physicists' jargon, the pile of sand/rice was in a self-organized critical state.

Buchanan shows that this kind of pattern is remarkably common in nature, from earthquakes to extinctions, to forest fires, to the effect of heat on magnets, to the way a frozen potato smashes. Only the steepness of the line on the graph varies. In the case of earthquakes, "double the energy [released by an earthquake] and an earthquake becomes four times as rare". In the case of fragments of frozen potato, "each time you reduce by two the weight of the pieces, their numbers increase by six". All of these relationships are 'power laws'.

So far, so fascinating — although I am not competent to judge whether Buchanan is getting the science right. What I can address is the question raised in his final six chapters: do such power laws exist in history? I can also address a question the author does not explicitly pose: if they do, so what?

The difficulty is that, splendid science writer though he is, Buchanan is no historian, and his argument begins to creak when he moves away from natural science. This is partly because — as his references make abundantly clear — his reading of history has been somewhat selective. More seriously, I think it is because he does not really understand what history is as a discipline.

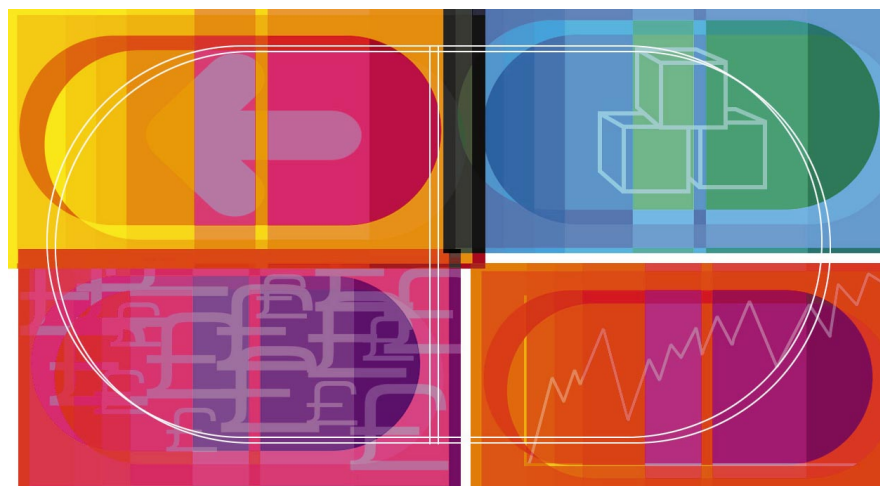
For Buchanan, it is an exciting possibility that versions of the 'power law' might exist for finance, war or politics. I have three difficulties with this. First, the quality of evidence is only good enough to sustain his argument in the case of financial markets (the data he uses for wars are just too unreliable). He needs, however, to read more: not only more of the work of Robert Shiller, but also Andrei Shleifer's research on irrationality in financial markets, to name only one of many economists who have worked on this. (Unlike most historians, economists can do maths, so this really is a big field.)

Second, the analogy between natural catastrophes and wars or revolutions — and it remains only an analogy — is not new. In the 1920s Winston Churchill likened international relations on the eve of the First World War to the interaction of planetary bodies: "If they got too near the lightnings would begin to flash, and beyond a certain point they might be attracted altogether from their orbits ... and draw each other into collision." David Lloyd George claimed the nations had "slid" into war — like grains of sand, perhaps. In other words, it was all inevitable. This was, of course, a convenient get-out for the politicians who had argued for war when the issue was decided on 2 August 1914.

My third quibble is that history doesn't gain much as a discipline even if it turns out that wars are just like earthquakes. Where do we go from here if, as Buchanan suggests, there really is a negative linear relationship between the size of financial crises or wars and their frequency, and if the same size of trigger can unleash anything from a 'correction' to a crash, or from a skirmish to a world war?

The author is rather scornful of traditional history, the sort that tries to apportion responsibility for the outbreak of the First World War ("The trouble began a week ago in the remote West, where in the early evening a single grain of sand fell on a portion of our pile that was already very steep"). But that is just what history is about. Ultimately, Buchanan has made the error of confusing history with social science. The reality, as the historian R. G. Collingwood pointed out long ago, is that history is our interpretation of past thoughts that happened to be written down or otherwise preserved. We do not really study the causes of wars, but what people at the time thought were the causes. And our aim in retrieving their thoughts is not so much to explain how things happened as to understand how they *seemed* to have happened.

If history is only a succession of economic, social, political and cultural earthquakes — usually small tremors, once in a blue moon really huge, with the timing more or less random — then we are indeed, as Shakespeare put it, "as flies to wanton boys". That



certainly delivers the *coup de grâce* to models of history based on regular cycles over time — and good riddance. What it does not do is make the historian's strange communion with the dead any more or less superfluous. ■

Niall Ferguson is at Jesus College, Oxford OX1 3DW, UK. His latest book, *The Cash Nexus: Money and Power in the Modern World*, will be published by Penguin next February.

From pioneers to marketeers

From Alchemy to IPO: The Business of Biotechnology

by Cynthia Robbins-Roth
Perseus: 2000. 272 pp. \$26, £15.95

Rino Rappuoli

Cynthia Robbins-Roth has written a real page-turner, as full of suspense as the most exciting novel. Yet *From Alchemy to IPO* is about the history and theory of the US biotechnology revolution. Robbins-Roth's enthusiasm for her subject stems from the fact that she was fortunate enough to share the early days with its pioneers, and her enthusiasm comes across on every page.

As we read of the early times at Genentech and the excitement of the scientists who carved out the biotech world, we also learn about the secrets of successful early biotech, and the mistakes that meant failure. The message that comes across throughout the book — and is even in the title — is that biotechnology was only partly a scientific revolution; an equally important revolution was in the financing of biotech businesses.

In the 'stagnant' world of that time, when only big pharmaceutical companies had the resources to fund drug development (which typically required some \$150 million and more than ten years before producing a return), two events changed the scene. First was the US law that in the late seventies legal-

ized the investment of pension funds in venture activities, thereby making large amounts of venture capital available. The second was the Genentech initial public offering (IPO) on 14 October 1980, which showed that slow and expensive drug development could be financed in the public market, without the need to rely on money from large drug companies.

Robbins-Roth's history of biotechnology focuses mainly on pioneers such as Genentech, and on today's most successful companies, such as Amgen. She makes clear that the companies that prospered best were those that from the outset hired managers who had experience in drug development with large pharmaceutical companies. Companies that allowed scientists to remain in charge for too long found it a struggle.

Another instructive history covered by Robbins-Roth, which shows how investors can make money even when a technology fails, is that of monoclonal antibodies. This was the technology that launched Hybritech in 1978. But in this case the first-generation technology was insufficiently advanced, and it was 20 years before a product was approved for human therapy.

Robbins-Roth covers the technologies that are biotechnology's building blocks: genomics, biochips, microarrays, antisense, gene therapy, tissue engineering, combinatorial chemistry and agricultural biotechnology. In an accurate and readable narrative that insiders and general readers alike will find interesting, she describes the steps involved in developing a drug for the market, including the associated risks — there is always something to learn.

Although the fundamentals of biotech are introduced briefly in the early chapters, later chapters lucidly describe the biotech business, product development and financing. Robbins-Roth provides a detailed guide on how to start a company: "you need an excellent management with relevant industry experience ... a strong technology platform to generate multiple product opportunities ... access to financing alternatives when

the public markets are closed ... the ability to articulate business opportunities, not just sexy science ... the right corporate location", and three key people: "An academic scientist ... an experienced pharma executive ... and a business/finance person with a track record." Robbins-Roth even provides a guide to all the financing steps taken by a start-up company as far as the IPO.

And last but not least, she includes a guide for investors. Again, her advice is sensible, and is supported by data in the form of tables and figures. While providing tips for reducing the risks of investing, Robbins-Roth also conveys the message that biotech is a solid and sustainable business, unlike many others that may appear more attractive in the short term but which are more susceptible to the caprices of the market. Importantly, she explains how the results of phase I and phase II clinical trials should be interpreted, and how to avoid panicking or overreacting.

Robbins-Roth is also a strong believer in the future of biotechnology, predicting that the tremendous progress of the past 20 years is just the beginning: "the 21st century will be the century of biotech." In a chapter entitled "Biotech star wars", she describes the future as predicted last year by NASA administrator Dan Goldin, who, dressed as a Jules Verne of the twenty-first century, described breathtaking trips in space beyond the solar systems. He talked of the colonization of other planets, missions guided by DNA-based computers with neural networks capable of adaptive learning and a billion times more efficient than today's silicon-based systems. During these long trips, biotechnology would enable food to be produced and could change a planet's unfriendly atmosphere to make it suitable for human colonization. Finally, through biotechnology, drugs could be developed to combat the unknown toxins encountered in space, and to produce regenerative and repairable body tissues. By the end of the chapter you just hope that you will live long enough to be able to see at least part of this future.

The book is both eminently readable and a source of useful data — as tables and appendices — that would be difficult to find in databases. I consider it by far the best book on biotechnology written so far. I recommend it to all biological sciences graduate students and their teachers, to managers and scientists in the biotech industry, and to academic researchers.

The book is also a must for people thinking about setting up a biotech company, and for those interested in investing in the biotech sector. I would even recommend it to people who don't invest in biotech, because it may convert them into investors.

Policy-makers, too, can learn from Robbins-Roth's message. This is that, although the innovation in biotechnology was the technology, a real impetus was given by the

financial environment, facilitated by appropriate laws, which created new ways of funding drug development.

Rino Rappuoli is at IRIS, Chiron SpA, Via Fiorentina 1, 53100 Siena, Italy.

Secrets will out

Battle of Wits: The Complete Story of Codebreaking in World War II

by Stephen Budiansky

Free Press/Viking: 2000. 448 pp. \$27.50/£20

Simon Singh

For decades, the code-breakers at Bletchley Park in the United Kingdom, who battled against the German Enigma cipher, were forbidden to talk about their work. It was only in the mid-1970s that the British government declassified Bletchley Park's role in the Second World War, and for the first time it became clear how code-breakers had made a series of brilliant scientific, mathematical and engineering breakthroughs that had influenced the course of the war.

Several books followed declassification, the first being *The Ultra Secret* by F. W. Winterbotham, who had been partly responsible for the government's change of policy. Other publications include *The Hut Six Story* by Gordon Welchman, one of the leading figures at Bletchley Park; *The Codebreakers*, a collection of essays by other Bletchley veterans; *Seizing the Enigma* by the distinguished cryptography historian David Kahn; and *Station X* by Michael Smith, which accompanied an excellent television documentary series of the same name.

These and other books were discussed last year in a Royal Society of Literature lecture entitled "Bletchley in Books", given by the eminent military historian John Keegan. After examining the various approaches of the many authors, he acknowledged that each book had its strengths, but his conclusion seemed to be that there was still room for the ultimate book about the Enigma story.



By writing *Battle of Wits*, Stephen Budiansky has taken up the challenge, daring to subtitle his book *The Complete Story of Codebreaking in World War II*. Indeed, Budiansky does not stop at the achievements of British code-breakers; he also attempts to tell the story of their US counterparts, who struggled against Japanese ciphers.

Budiansky, a mathematics graduate and a science writer on diverse topics, begins by explaining how the First World War renewed American interest in military cryptography. When the United States entered the war in April 1917, it was the 27-year-old hard-drinking, womanizing gambler Herbert O. Yardley who persuaded the war department to create a clandestine cipher bureau, which he then commanded.

Yardley maintained his so-called Black Chamber by setting up a company to act as a front, which was secretly subsidized by the state and war departments.

After the war, he demonstrated that he was still worth funding by cracking the Japanese diplomatic codes and reading their bargaining strategy for the 1921 Washington Naval Conference, which had been convened to agree limits on shipping. The Japanese declared publicly that they would settle for nothing less than 7 tons of their own shipping for every 10 tons of British and US shipping, but Yardley knew that their fall-back position was 6.5 tons. He advised the Americans and British to refuse 7 tons, and indeed the Japanese caved in and accepted 6.5 tons.

In 1924, the new secretary of state Henry L. Stimson accidentally learned of Yardley's activities and closed down the Black Chamber, later stating that "Gentlemen do not read each other's mail". Yardley subsequently embarrassed his former bosses by publishing his memoirs in *The American Black Chamber*, a notorious bestseller.

Despite Stimson's efforts, code-breaking secretly continued under the leadership of William Friedman, America's foremost cryptographer of the war years. His achievements ranged from his pioneering work on the polyalphabetic cipher to the development of a highly efficient randomization process for encryption. This involved throwing a stack of encoding cards in front of a large fan and then picking them up off the floor.

In parallel with the US history, Budiansky describes the development of British code-breaking. The result is a complex narrative that roams across time and space. However, the pay-off is the ability to compare and contrast the code-breaking efforts of the two nations. For example, whereas the British developed electro-mechanical code-breaking devices known as bombes, the Americans relied heavily on well-established IBM punch-card technology. At the end of the nineteenth century, the American engineer and inventor Herman Hollerith had built

machines that combined the punch-card format for storing data with an automatic sorter and tabulator. The idea took off after 1890, when the Census Bureau hired 50 machines to analyse census data. Soon after, punch-card machines were used by a variety of businesses to keep track of their operations. In 1911, Hollerith's Tabulating Machine Company merged with two other firms and became International Business Machines.

Friedman's request for punch-card machines to help break codes was initially refused, even though they were already being used by the military for accounting. Eventually, a newly appointed quartermaster-general, who was more used to traditional accounting techniques, stopped using the machines assigned to him, and Friedman was able to have the lease signed over to his code-breaking department.

Many of the encryption systems Friedman encountered were enciphered codes. First, each word would be encoded as a five-digit number. But this alone is not secure, because frequently occurring number groups can be associated with frequently occurring words. Hence, the Japanese would encipher each number by adding another five-digit number from an effectively random list. The IBM machines could check hundreds of messages looking for repetitions, which indicated the rare occasions when the same encoded word had been enciphered in the same way. Such repetitions were the only way to crack enciphered codes.

Because so much has already been written about Bletchley Park, many readers may find the American sections of the book more interesting. It has been well documented how Bletchley Park gave the Allies an advantage in every arena from the Atlantic to North Africa, but much less has been written about the code-breakers at Arlington Hall — a former finishing-school for girls near Washington — who changed the course of the war in the Pacific.

In 1942, American code-breakers, who had cracked the JN-25 cipher, were able to warn of the planned Japanese attack on America's Midway Islands, which meant that the US Navy was prepared. As a result, the Japanese Navy suffered its first defeat in three centuries. The following year, another code-breaking coup allowed American fighter planes to intercept and assassinate Admiral Yamamoto, commander-in-chief of the Japanese Fleet.

It will be interesting to see how librarians catalogue Budiansky's book. Although he does discuss the technology and science involved, *Battle of Wits* contains long sections of well-written and fascinating military history. Browsers who discover the book on the history shelves might be surprised to find scientific explanations that can be quite intense. Those who find the book in the sci-

ence section may be a little disappointed by the lack of scientific emphasis, but they should relish the excellent historical detail that puts the work of Arlington Hall and Bletchley Park into context.

The story of cryptography in the Second World War is one of the great scientific tales of the twentieth century. Budiansky has succeeded in telling it with enthusiasm and insight, delivering a book with style and substance. ■

Simon Singh is a science journalist and author of Fermat's Last Theorem, The Code Book and The Science of Secrecy.

Now you see it, now you don't

The Undergrowth of Science: Delusion, Self-deception and Human Frailty

by Walter Gratzer

Oxford University Press: 2000. 338 pp.
£18.99, \$27.50

Douglas R. O. Morrison

It came as a shock to many when Richard Feynman reviewed a subject that was backed by two opposing theories, each with strong experimental support, and concluded that the only answer was that one of the experiments must be wrong. He was right, as usual, but for those educated in the Rutherford tradition of believing that experiments are the true basis of science, this was almost a heresy. We are accustomed to theories being wild hypotheses, but not physical experiments! Once one accepts this possibility, however, false results can be found to occur fairly frequently and we should take them into account.

The first person to study what happens when a scientist declares a wrong result, and others follow him, was Irving Langmuir, who, in a famous lecture, gave it the name 'pathological science'. He described cases such as N-rays, 'discovered' by a respected physicist, René Blondlot. N-rays collapsed after a single dramatic episode when the physicist Robert W. Wood visited Blondlot and was present when, with incredible accuracy, he measured the dispersion of the rays by a prism. When, in the darkened room, Wood removed the prism, Blondlot's measurements were unaffected. Wood published and effectively killed N-rays. Thanks to Langmuir, if one wants to hint that a series of results may be wrong despite confirmatory publications, a mention of N-rays gives the connecting clue.

When Martin Fleischmann and Stanley Pons published evidence for cold fusion and their findings were then supported by independent scientists, we talked of Langmuir



and pathological science. But this time there was no Robert Wood to strike the killer blow. If only Dick Feynman had lived longer and been able to use his wit on cold fusion.

Walter Gratzer has written a book describing many examples of pathological science. He is a good writer and holds the reader's attention well. The first part of the book takes the major case studies given in Langmuir's lecture, and expands on them using hindsight. He adds several other examples, including the more recent case of cold fusion. But his strongest criticisms are of the medical profession, which he knows best. It seems incredible that a doctor or group of doctors could believe in a treatment and carry it out for many years without people remarking that it was not acting as a cure but was instead killing many patients.

One example is ptosis, which means 'dropped organs'. Operations on the kidney began around 1883 and the death rate from removing a dropped kidney was about 50%. Later operations sewed the kidney to the abdominal wall. Many thousands of these useless operations were performed over some 50 years, to correct a condition that caused no deaths. Removing the colon was recommended at one time despite the 18% death rate of the operation. Gratzer also describes and attacks homeopathy, although this is still popular, particularly in France.

The second half of the book extends these essentially scientific case studies to the influence of politics and racial prejudice on science. The deadful story of how Trofim Lysenko destroyed genetics in the Soviet Union is well known but worth re-reading. Gratzer also describes how the same procedure of having well-organized meetings at which possible opponents of the establishment are denounced and then removed, was prepared for a group of physicists. Just before the first such meeting, in 1949, Igor Kurchatov, the father of the Soviet atom bomb, saved them by declaring to Lavrenti Beria, head of the secret police, that quantum mechanics and relativity theories were essential for developing an atom bomb. According to Gratzer, Beria consulted Stalin, who, speaking of the physicists, declared:

"Leave them in peace; we can always shoot them all later."

The important lesson is that science is universal, and when it is viewed as Soviet science or Aryan science, disaster can often be the outcome. The science of eugenics, literally 'good birth', assumed that all human characteristics, intellectual and moral as well as physical, were inherited, and hence people of 'inferior' races or those with hereditary diseases should be controlled.

The concept of Aryan science started long before Hitler came to power and was based on the principle that the Aryans were a superior race. The Nazi government first applied these ideas to its own people, and some 300,000 people considered physically or mentally inferior were sterilized between 1933 and 1939. After 1936, killings by doctors in hospitals and asylums began. In 1939 euthanasia was legalized to replace sterilization, and finally, no fewer than 70,000 patients in institutions were secretly killed. Later, this belief in racial superiority led to the killing of millions of Jews, Poles, gypsies and others in concentration camps.

Eugenics started in Britain in 1869 and quickly spread to the United States, where eugenic measures such as marriage laws preventing unsuitable unions were introduced. Compulsory sterilization followed in many states, starting with Indiana in 1907 and lasting well into the sixties. Other countries followed suit. In Sweden, compulsory sterilization became legal in 1934, and over the next 30 years about 1% of the population — 63,000 Swedes — were sterilized for reasons of race or social undesirability. Eugenics is now almost dead, but racial discrimination is alive, although usually bearing other names, such as 'ethnic cleansing'. The fight against eugenics continues.

Gratzer writes as a historian, and so his book lacks the charm of personal involvement found in Langmuir's contributions — Langmuir played the part of Robert Wood by demonstrating to experimental physicists Bergen Davis and Arthur Barnes that they were counting imaginary scintillations. Gratzer could have discussed cold fusion in more depth, as this controversy continues today. In fact, the true believers held a meeting in Italy last May which was sponsored by the three main Italian official research organizations.

For a wrong result to be believed and for the idea to spread, the reputation of the people involved is very important. Bob Park, the author of a somewhat similar book, *Voodoo Science*, says that, no matter how crazy the claim, it is always possible to find a physicist with a PhD to support it. Reading Gratzer's book, one is tempted to say that there is a 50% chance that a Nobel laureate will support the claim.

Interestingly, Gratzer shows that scientists who make bad errors tend to be treated

kindly by their colleagues. This is probably because the profession believes in self-regulation, and possibly also because there is a feeling of 'there but for the grace of God, go I'.

What will be the next example of pathological science, for there are surely many more still to come? Possibly it will be something we all desire — a new energy source, life on Mars. I would recommend Gratzer's book as a tool to help us recognize it sooner and fight it effectively. ■

Douglas R. O. Morrison is an honorary staff member of CERN, CH-1211 Geneva 23, Switzerland.

Funny, I thought this was science

Laughter: A Scientific Investigation

by Robert R. Provine
*Viking/Faber & Faber: 2000. 256 pp.
 \$24.95, £12.99*

Steve Blinkhorn

If a lion were laughing, how could we tell? Or is the very idea nonsense? It is easy enough to suppose that early hominids felt fear, despair, elation, sorrow — but what would count as a simian snigger, and does joviality depend on a joke? Laughter has an awkward place in any account of emotion, perhaps because of the suspicion that it is inextricably linked with self-conscious cognitive processes and so is the preserve of our own species.

There is a craft of humour, whose aim is to provoke laughter. But just as the study of paths is not exhausted by an examination of tragic literature, there is more to laughter than the creation of comedy. This "scientific investigation" of laughter strives to remain on the subject and to avoid being sidetracked into the theory and practice of humour, but

in the process makes some exceedingly bad jokes and some very dubious claims. On the other hand, it has the conspicuous merit of at least tackling its subject from many points of view, however unpersuasively. Journalistic in style, it starts "with the 'who', 'what', 'when', and 'where', of laughter". The fact that the 'how' and the 'why' are missing is the clue to the essentially ethological, tractarian nature of the enterprise. "Laughter, like speech, is a vocal signal that we seldom send unless there is an audience." But how can you *know* that? Is this not just one author's perception? It may be a sign of serious mental weakness, but let me confess to laughing out loud at books, and broadcast comedy, and — yes — even the works of Daedalus in the pages of *Nature*, when no one else is present.

Are these vicariously social situations? Very likely, but then the notion of an audience becomes vacuous. Provine had volunteers record in diaries their solitary behaviour, and not surprisingly they laughed more in company than alone. Most people drink more alcohol in company than alone, and while that surely tells us something about the prominence of social drinking in Western society, it tells us nothing about the chemistry and metabolism of ethanol.

Worse still, so far as solitary vice is concerned, in the days when I wrote comedy for performance, I knew that if it didn't make me laugh out loud as I wrote it, it would leave an audience cold. Of course, that is to fall into the trap of equating laughter and comedy, and whereas comedy is culturally conditioned, ticklishness isn't. Unfortunately, because a great deal of consensual tickling and concomitant laughter takes place in circumstances beyond the ethologist's gaze, a comprehensive account based on frequency, rhythm, intensity, location and mutuality must await some future Masters and Johnson (of titillation in the literal sense). Provine notes, somewhat sadly, that opportunities for engaging in this often highly ambiguous-



ly enjoyable activity decline with advancing years, as does recorded pleasure in the experience when it does occur. But perhaps willingness to participate enthusiastically in the business of rating, on a 10-point scale, the pleasure experienced by being tickled also declines markedly with age.

The irritating thing about what is, in principle, an attractive scholarly enterprise, is the sheer unevenness of the treatment. One might almost think it an unhappy collaborative effort. After some truly dire attempts at humour ("premature ejokulation", "laftus interruptus"), the reader is pleasantly surprised by stretches of good, pacy exposition of plausible science and intriguing insights from primatology and the study of autism. But then Chapter 9, "Laughing Your Way to Health", feels like an editorial imposition, and comes to no worthwhile conclusions at all. The following chapter, "Ten Tips for Increasing Laughter", solemnly advises the reader to "stage social events" and "provide humorous materials". We need a neurobiologist for this kind of advice?

What bothers me most is that a professor who, presumably, has spent many hours on his feet engaging the attention of eager youth on matters scientific feels it worth proposing that, as a public speaker evokes laughter from an audience, "*the brains of speaker and audience are locked into a dual-processing mode*" (author's italics). Classical manuals of rhetoric have more insight to offer. "Laughter is about relationships" — but only in the sense that Life Is About Relationships, a sense that does little to inform and nothing to explain.

Humour can be very culture-specific: recognition laughter is comprehensible only in terms of a set of expectations and experiences, the humour of incongruity only in terms of what would count as congruent, and neither yields much to this analysis. As for irony, well, that is perhaps in any case a peculiarly British taste, and possibly one of the great barriers to shared laughter between nations. I came to wonder eventually just

how much the author's sense of humour has been sidetracked by his professional interest in laughter. One is left with the feeling that, in his view, laughter is funny peculiar rather than funny ha-ha, and that putting this book together was rather less fun than he would like us to believe. ■

Steve Blinkhorn is at Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans AL1 3HT, UK.

The gene is dead; long live the gene

The Century of the Gene

by Evelyn Fox Keller

Harvard University Press: 2000. 192 pp.
\$22.95, £15.95

Jerry A. Coyne

Gregor Mendel's work was rediscovered in 1900 and Wilhelm Johannsen coined the word 'gene' in 1909. Since then, genetics has progressed from T. H. Morgan's work on the fruitfly *Drosophila* to the genome projects of today. In retrospect, it seems appropriate to dub the twentieth century, at least in scientific terms, 'the century of the gene'. But despite the title of her book, Evelyn Fox Keller disagrees.

The Century of the Gene is, in fact, a jihad against our notion of the gene. Keller insists that the gene is neither the stable, self-replicating entity we thought it was, nor a repository of information about development. To Keller, 'gene' is simply an outmoded term, a semantic straitjacket signifying something that can't be defined. Were she less constrained by publishing convention, I suspect her book would have been called *The Century of that Nebulous, Ill-Defined Entity Formerly Known as 'The Gene'*.

Keller, a philosopher and historian of science, is best known for *A Feeling for the Organism* (W. H. Freeman, 1983), her biog-

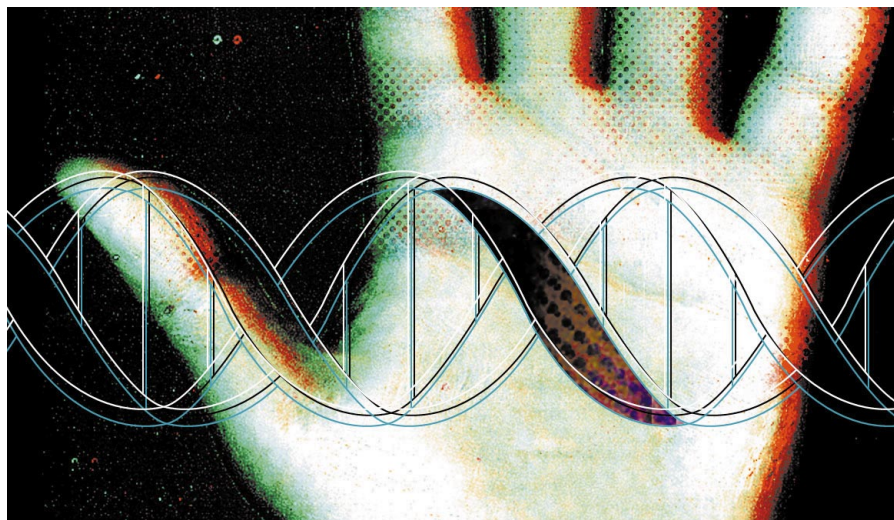
raphy of the geneticist Barbara McClintock, which was written for a general audience. Given the high technical level of discussion, *The Century of the Gene* is, however, clearly aimed at professional biologists.

Unfortunately, the book is long on complaint and short on substance, and ultimately fails to make its case against the primacy of the gene. Despite her repeated claims that the recent history of genetics is replete with "major reversals", "serious provocations" and "radical modifications", the gene emerges unscathed. Many of the alleged problems highlighted by Keller turn out to be semantic issues likely to be of little interest to either working biologists or serious philosophers of science. Moreover, the level of analysis is disturbingly superficial: Keller seems more interested in forcing genetics into the Procrustean bed of her thesis than in presenting a balanced argument.

She claims, for example, that the idea of the gene as a unit of structure or function is outmoded because some bits of DNA do not produce proteins, but instead regulate genes, because some genes can be spliced or read in alternative ways, and because the products of some genes perform several functions. Although it is true that genes are often complex, the word gene is still a perfectly good working term for biologists, especially when defined as a piece of DNA that is translated into messenger RNA. Farmers are still called farmers even though their job is far more complex than that of their predecessors.

Keller asserts that DNA is not a 'self-replicating' molecule because enzymes are needed for replication. She also claims that genes do not direct development because gene activation depends on many different factors (such as chromatin structure, egg cytoplasm and local differences in the cellular environment which turn on different genes in different tissues). Again, these are pseudo-problems: replication enzymes and many inducers of development are themselves products of genes. One might as well argue that political candidates are not self-promoting because they hire others to do that job for them. Certainly, non-genetic factors influence development, but ultimately we differ from chimpanzees because of our genes, not our environments.

The supposed non-autonomy and complexity of genes lead Keller to suggest that we should replace a reductionist approach to genetics with a more holistic programme that incorporates trendy concepts such as developmental networks and self-organization. But she does not specify how this approach would work. In fact, history shows clearly that the greatest triumphs of genetics have been born of reductionism: progress nearly always comes by first studying single genes and then examining their interactions with others. The remarkable advances in understanding the developmental genetics



of *Drosophila*, for example, confirm the value of reductionism in molecular biology.

An example of Keller's one-sided treatment of more substantive issues is her discussion of 'evolvability'. A recent buzzword in evolutionary genetics, evolvability refers to the idea that, in some species, natural selection may favour traits that increase the likelihood of future evolution. There is considerable controversy about whether and how this could occur, but Keller ignores these disputes. Instead, she promotes a particular form of evolvability that, she claims, is both ubiquitous and a radical challenge to modern Darwinian theory. She is wrong on both counts.

Keller argues that species have evolved ways of increasing their mutation rates to generate genetic variation — the raw material for further evolution — and that this evolution undermines the idea that genes are stable. Her evidence for 'adaptive mutability' is the observation that, in some microorganisms, various forms of environmental stress (such as starvation, ultraviolet light or extreme temperature) appear to activate genetic systems that increase the mutation rate. Although most mutations are harmful, some may be useful, and genetic linkage between 'mutator genes' and their adaptive products may drive mutators to high frequencies. Permanently increasing the output of new variants could accelerate future evolution. There are, however, serious problems with this argument.

Keller's prime example of adaptive mutability is the SOS repair system, a mechanism for DNA repair best characterized in the bacterium *Escherichia coli*. When pervasive, stress-induced damage overwhelms normal repair mechanisms, the SOS system comes into play. This system reverses many mutations, but in so doing introduces a few others. Keller suggests, as do some microbiologists, that the SOS system is an adaptation for increasing the mutation rate under stress. But as this system acts to repair mutations, a more parsimonious explanation is that it evolved simply as a second line of defence against DNA damage and, like many adaptations, is imperfect.

Unfortunately, Keller mentions neither this alternative explanation nor the continuing debate about the nature and meaning of stress-induced mutability. Moreover, she fails to note that selection for higher mutation rates via linkage does not work in sexually reproducing organisms. In such cases mutator genes will be separated from their adaptive products by recombination and then eliminated by natural selection.

Finally, such inducible mutator systems can yield an adaptive response only to factors that impinge directly on DNA molecules. In multicellular organisms with separate germ cells, most forms of selection do not work this way. The presence of lions on the

savanna does not increase the mutation rates in gazelles.

Some individual genes, including vertebrate antibodies, have apparently evolved new ways of generating variation as an adaptive response to constantly changing selection. But this, as well as any selection for inducible mutation in bacteria, can be completely explained by evolutionary genetics. By unwarranted extrapolation from bacteria to all organisms, Keller grossly exaggerates the challenge of evolvability to both Darwinism and genetics.

Keller concludes that "gene talk", the argot of geneticists, is *passé* because of "accumulating inadequacies of an existing lexicon in the face of new experimental findings". Gene talk persists, she says, because it is an easy way for biologists to communicate, and because it helps geneticists get grants and biotechnology companies make profits. Her remedy is to call for a new vocabulary that incorporates concepts from engineering and computer science. Sadly, she fails to suggest what words or concepts we need. Although my enthusiasm for neologisms is limited, they can be useful, as in physicists' distinction between 'mass' and 'weight'. But the notion that geneticists are semantically challenged is simply silly. There is not the slightest evidence that future advances in genetics will be stalled by an outmoded lexicon. What we need is more work, not more words.

The physicist Richard Feynman, famous for his one-liners, supposedly said that the philosophy of science is as useful to scientists as ornithology is to birds. His criticism is overstated, because philosophy can give scientists intellectual perspective on their work. *The Century of the Gene*, however, ranks as opinionated and poorly informed ornithology. The gene is no albatross. ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

With a hammer and passion

Trilobite! Eyewitness to Evolution
by Richard Fortey

Knopf/HarperCollins: 2000. 288 pp.
\$26/£15.99

Philippe Janvier

Palaeontology is one of the rare areas of science that teenagers can tackle by themselves, especially if they live in a fossil-rich area. All that is needed is a hammer and passion. Such an early training certainly helps the happy few who finally manage to become professionals, as was the case for Richard Fortey. I empathize with him, as I too had a youthful passion for fossils.



In this book, Fortey gives a passionate and often lyrical account of his life with trilobites, a group of extinct marine arthropods related to the living horseshoe crabs and spiders, which lived from 540 to 260 million years ago. Trilobites are indeed fascinating. They look a bit like large woodlice, but show an amazing diversity of morphologies. What's more, their preserved anatomy is complex enough to allow scientists to reconstruct the group's relationships and evolutionary history.

In a vivid, popular style, full of didactic metaphors and anecdotes, Fortey writes about how his passion for trilobites arose when he was 14, how he was trained by his mentor Harry Whittington, how he discovered new trilobites in the rocks of Spitsbergen, China, Thailand and Australia, and his life with colleagues in the small community of trilobite specialists. He also recounts the history of trilobite research, how early palaeontologists gradually revealed the most intimate details of trilobite anatomy: the amazing structure of their eyes, and their long-elusive appendages and gills. Fortey uses trilobites to explain how palaeontologists work, from basic fieldwork and identifying and describing species to far-reaching generalizations about evolution. Seen in this way, the book is an excellent introduction to the basic practice of palaeontology and systematics.

Trilobites have been in at the birth of several theories about the process of evolution. For example, there is Niles Eldredge and Stephen Jay Gould's 'punctuated equilibria', an evolutionary pattern where a species shows a long period of stability (equilibrium) and is then suddenly replaced by its closest related species (punctuation). Or there is McNamara's 'heterochronism', an evolutionary process involving shifts in the timing of the development of certain organs, and hence shifts in the morphology of the entire organism. Trilobites are also there at the Cambrian explosion — the period 540 million years ago where most major animal groups appear suddenly in the fossil record. Fortey uses trilobite examples

to explain and discuss these palaeontological classics.

Here and there Fortey depicts the world of trilobite specialists, past and present, and recounts his field adventures in a 'Tintin-and-the-lost-trilobite' style. Those who know the field or palaeontologists may, however, dislike the way he uses examples of how rude and sectarian some in this — otherwise humble and idyllic — community can be in settling scores with certain colleagues and journal editors.

Fortey's youthful enthusiasm is refreshing, but unfortunately this is often what managers of science dislike, as if enthusiasm and obsessive passion were a sin, a childish reaction of amateurs. Those who decide who will be selected for a life in science tend to prefer the cold-blooded ones, who have only scholarly experience filtered by the academic world. Yes, Fortey was very fortunate to be able to have such a straight career in trilobites, despite his passion. But some may say that his enthusiasm hinders him from taking a cold look at what trilobites can really tell us. He sees them swimming in the Ordovician sea, he imagines what they could see through the numerous calcite lenses of their eyes.

This book is pervaded by the idea that time is something very important, and unique, in palaeontological data. The author seems to be able to read evolution through the rock layers. He apparently expects something new and odd from his trilobites just because they are old, but at the same time he points out repeatedly how perfect they are in many respects, and their striking analogies to modern arthropods. To him, "palaeontology is a daughter of Old Time, or it is nothing", and time "tests the durability of scientific revelation".

Others see fossils as mere bearers of new characters or associations of characters, whose position in the tree of life sometimes reveals the pattern of relationships of other, distantly related living organisms. In this view, time is no longer a property of fossils, which are simply new organisms added to the diversity of life and ranked according to the logic of character distribution.

Trilobites are thrown clear of the Cambrian explosion and, as a figure caption says: "where are the ancestors of the trilobites?" Fortey admits the power of cladistics and character analysis in such cases, when time is mute and only characters speak. He also insists on the extraordinary diversity of the trilobites, which he portrays as a fanciful parade. But one would have expected some orderliness to emerge from this parade and for the genealogy of the group to be outlined.

As a whole, the book is a sprightly description of everyday life in palaeontology and of a small scientific community throughout time, with its great names (such

as the French-born Joachim Barrande, to whom Bohemia was only a place of exile — not a "home land"), its victims, tragedies, glories, awkward customers, (presumed) fraudsters, and many ordinary workers who accumulate data and new insights. Amateur palaeontologists, as well as professionals, who like nature's wonders and evolutionary scenarios will certainly love this book, and students will discover the kind of world they might step into.

Philippe Janvier is in the Laboratoire de Paléontologie, Muséum National d'Histoire Naturelle, 8 rue Buffon, 75005 Paris, France

A diamond among dinosaurs

Tyrannosaurus Sue: The Extraordinary Saga of the Largest, Most Fought Over T-Rex Ever Found

by Steve Fiffer

Freeman: 2000. 248 pp. \$24.95, £16.95

Douglas Palmer

To be selected as the main attraction at Disney's Animal Kingdom adventure park is a big enough achievement for any long-dead animal. Nevertheless, a fine specimen of *Tyrannosaurus rex* has gone one (indeed, two) better. It has the ultimate accolade of having been chosen as a marketing superstar for McDonald's and as the 'Hope Diamond' — the main crowd-pulling attraction — of dinosaurs for Chicago's Field Museum. Now that is fame, even if it has taken more than 65 million years to achieve it.

Mind you, this is no ordinary *T. rex*, but, as Steve Fiffer claims, "what many call the most spectacular dinosaur discovery to date". And with a price tag of \$8.36 million, Sue is probably the most valuable fossil ever found.

The dinosaur was discovered in 1990 by

Sue Hendrickson, who was working with Peter Larson at the Black Hills Institute of Geological Research, a private company of commercial fossil hunters. Larson paid Native American rancher Maurice Williams, on whose land 'Sue' was found, \$5,000 for what has turned out to be the most complete and best-preserved specimen of a *T. rex* known so far.

In May 1992, armed FBI agents, sheriff's officers and national guardsmen took Sue out of Larson's hands and put her in protective custody. The previous month Gregg Bourland, the chairman of the Cheyenne River Sioux, was reported to have agreed to a partnership with Williams to determine the future of the dinosaur. They were claiming that the dinosaur had been illegally taken from within the boundaries of their reservation, within which Williams's ranch lay — a charge of which Larson was eventually acquitted. And Williams was claiming that he had not sold the specimen to Larson but "only allowed (him) to remove it and clean it and prepare it for sale".

In 1995 Larson went on trial for knowingly collecting specimens from public lands. He was acquitted of taking Sue from public lands and only convicted of two out of 33 charges, and these were customs violations, but was sentenced to two years in prison and two years' parole. He was released after serving 18 months, just six weeks before Sue went up for auction at Sotheby's in New York.

All good, clean fun? Far from it. As Fiffer so ably shows, with millions of dollars involved, the fighting over Sue has been anything but clean or fun — and he should know, he's a lawyer himself.

David Redden, the executive vice-president of Sotheby's in New York, had been following the custody battle for Sue with the interest of a predatory carnosaur. "It was a wonderful snapshot of many aspects of our social compact," says Redden. "The rights of palaeontologists, Native Americans, the



rights of the public." But for him the real challenge was "how to move it to the next owner" — something he felt sure Sotheby's could assist with. Indeed it did, and did very nicely in commission, too. Sue has now gone to the Field Museum in Chicago, where she joins one of the world's great natural history collections: a treasure trove of more than 20 million specimens, including two very large and fine sauropod dinosaurs, one of which has been pulling in the punters for nearly 100 years.

Dinosaur icons of 'mass and might' suited the mid-nineteenth century, when they were first discovered and exhibited, to public acclaim, at the Crystal Palace in London. Nowadays, though, 'lean and mean' is the name of the game. That coincides with the public perception of *T. rex*, even if it is not scientifically accurate. As Fiffer puts it, *T. rex* has become a must-have for a museum, something that the Field Museum's president John McCarter also appreciates. "You need a big, defining exhibit," McCarter explains. "The most visited object [in a museum is] the Hope Diamond at the Smithsonian. It brings them in."

But there's a more important reason behind this. "The reason we do dinosaurs," says McCarter, "is so that we can do fish." The big crowd-pullers generate revenue to support less glamorous endeavours, which may have more scientific value. But the Field was not prepared to spend a packet on a dinosaur, no matter how spectacular, unless it had some intrinsic scientific value. This Sue does have: her relative completeness will help to resolve some of the outstanding questions about tyrannosaur skeletal anatomy.

Fiffer expertly tells the fascinating story of behind-the-scenes networking, wheeling and dealing at which many of the main players such as McCarter were so proficient. McDonald's was lured into the game by the possibility of active participation in the project rather than just a donor plaque. As luck or fate would have it, McDonald's had just entered a ten-year marketing deal with Disney and was sponsoring Disney's Animal Kingdom adventure park in Orlando, Florida. So Disney was in and the corporate package fell serendipitously into place once they had won at the auction. The problem was that Sue had already achieved considerable trophy status, making it virtually impossible to guesstimate what her price would be.

Tyrannosaurus Sue contains all the ingredients for a good story — the Wild West, guns, romance, the Sioux, drama, jail, dollars and not just any dinosaur, but *T. rex* itself. What's more, it is all true. ■

Douglas Palmer is at 31 Mawson Road, Cambridge CB1 2DZ, UK. His book *Neanderthal* was published by Channel 4 Books last month.

Chemists in their element

Mendeleev's Dream: The Quest for the Elements

by Paul Strathern

Hamish Hamilton: 2000. 309 pp.

£12.99

Gautam R. Desiraju

Paul Strathern is more philosopher than chemist, and in this engaging book he has set the life work of Dmitri Mendeleev against an appropriately contemplative backdrop that stretches back in time and is replete with connections linking science, philosophy and human nature.

The quest for the elements begins with the origin of scientific thought — Thales of Miletus in Asia Minor and his successors, Socrates, Plato and Aristotle. We stumble through the dark ages of alchemy with Zosimos, Jabir ibn Haiyan (Geber), Avicenna and Paracelsus, even as glimmers of rationality are perceived in the form of Roger Bacon, Nicholas of Cusa and Giordano Bruno. An excursion into the development of astronomy and physics in the Renaissance follows, but as Strathern aptly states, people had to understand what science was through these quantitative sciences before the more qualitative science, chemistry, could emerge.

In any event, physics and chemistry were not well differentiated when Robert Boyle provided an embryonic definition of a chemical element. Despite all this meandering, the stage is set quickly by Carl Wilhelm Scheele, Antoine Lavoisier, John Dalton, Jöns Jacob Berzelius and others, with the discovery of several elements. This is described in a charming chapter, "Things Never Seen Before", and the drama heightens. The climax approaches quickly, via Johann Döbereiner, John Newlands and the esoteric Alexandre-Émile Beguyer de Chancourtois, and finally the protagonist appears in all his glory.

So much for the narrative. The appealing feature of the book, however, is its philosophical content. Strathern refers constantly to the thinkers and the doers, and this is but analysis and synthesis. Indeed, chemistry is all about analysis and synthesis, not just of substances but also of thought. Success in the quest for the chemical elements always depended on an optimal interplay between these two great philosophical elements. In both chemistry and philosophy, analysis always presupposes synthesis, but in turn always leads to more and better synthesis. So, in hindsight, it is easy to understand why both the ancient Greeks, who attempted analysis without synthesis, and the alchemists, who believed they could sustain synthesis without analysis, were doomed to failure.



What, then, of Mendeleev? Analysis and synthesis came together intuitively, to his quicksilver mind, resulting in his *magnum opus*, the periodic table. The main challenge in cracking the code of the elements was analytical rather than synthetic. Even if one accepted Lavoisier's basically correct definition of an element, one was never fully certain if a particular chemical entity was an element or not, and even then, one did not know its properties accurately enough. All European intellectuals of the time were aware of Immanuel Kant's distinction between things as they appear and things as they are; recognizing what one does not know is the cornerstone of this philosophy. However, Mendeleev, with his Tartar origins and encyclopaedic knowledge of chemistry, went strategically and daringly further. He assumed that what was unknown conformed to what was known: gallium and germanium were the missing cards in his game of patience. Surely, the fault line between occidental and oriental philosophy lies somewhere in the icy wastes between Königsberg and St Petersburg.

Although the book is full of detail, it always remains pleasantly readable. Many a little-known vignette is presented with irony and dry humour, be it Isaac Newton's alchemical labours and misogyny, Johann Wolfgang von Goethe's amateurish scientific dabbings or Henry Cavendish's eccentricities.

In all this, though, there is room for introspection. Will future generations smile at patronizingly at our foibles as we do when we are told of the hare-brained scheme of Hennig Brand and Gottfried Leibniz (of calculus fame) to transport vast amounts of human urine, obtained conveniently from beer-guzzling miners in the Harz mountains, in horse-drawn carts over 100 kilometres of rutted roads to Hanover, all in the name of obtaining elemental phosphorus for undefined and unclear purposes? Maybe so, for scientific activity is a long-standing business. Every scientific battle won is a battle lost and, in the end, scientists are just first observers of a grander design. All this and

more can be gleaned from this informally written book, which is well recommended to the general reader. ■

Gautam R. Desiraju is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

A Saturnine binge

Brush with Death: A Social History of Lead Poisoning

by Christian Warren

Johns Hopkins University Press: 2000. 384 pp. \$45, £35

Thomas Dormandy

"America is being lead poisoned," Christian Warren declares in the first sentence of this book. This is, of course, nonsense. There are several environmental agents that harm and shorten the life of the average American today, but lead is no longer in the top league.

Yet for 20 years it made the headlines. It was in the late 1960s and early 1970s that a series of well-conducted surveys first revealed that the blood lead of the average American had risen to a level 100 times higher than had long been accepted as safe, though not necessarily desirable. Impassioned battles had already been fought to eliminate lead poisoning from the lead-using industries and to combat the terrible 'silent epidemic' of childhood plumbism, which induces intractable mental and behavioural problems. But the new menace was both of a different kind and of a different order.

For thousands of years lead had been one of the most useful as well as one of the most versatile of metals, and the discovery that one of its organic derivatives, tetraethyl lead, should also prove to be the cheapest and most effective antiknock agent in petrol came as no surprise. Nor was it surprising that, as the number of motor cars on American roads rocketed and the monstrous power of their engines became one of their selling points, a certain amount of the additive should escape into the atmosphere. Once in the atmosphere, its inhalation by all who shared their air space with cars was inevitable.

Nor was the first reaction to the rising level of lead in people's blood one of alarm. The doctrine propagated by the lead-using industries had been summarized by Robert Kehoe in the 1960s. "The natural environment of the planet is such," Kehoe had said, "that the occurrence of lead in the tissues, body fluids and excreta of its human inhabitants is inevitable. Fortunately, it appears that an equilibrium is quickly established ... whereby the stream of lead absorbed is balanced by a counterstream of lead issuing

forth from the tissues of the body via the excretory routes. ... No doubt that equilibrium can be temporarily disturbed ... but an essential balance over the life span of an individual is maintained."

This idea of self-regulation was deeply reassuring. Indeed, in the 1960s and 1970s few people were concerned about rising blood-lead levels. But by the 1980s public response to environmental issues had undergone a sea-change. Was this related to the crumbling of communism and, with it, the fear of nuclear war as the great, all-pervasive menace? Contrary to the inane phrase fashionable at the time, it did not signal the end of history. Almost at once, other dangers began to fill the anxiety vacuum. Suddenly, there were the ozone layer, carbon monoxide emissions, passive smoking, the ever-growing mountain of non-biodegradable trash, radioactive waste. And to cap it all, there was now lead at the very heart of the 'world's greatest democracy'.

Coming at the right moment, dubious historical insights can ignite public imagination more readily than mountainous statistics. In *Rome's Ruin by Lead Poison*, S. Colum Gilfillan suggested that plumbism "is to be reckoned as the major influence in the ruin of Roman culture". The most important sources in the antique world were the growing use of lead as a preservative of wine, lead glazes and lead pipes. Gilfillan's thesis (which had somehow escaped the vigilance of Edward Gibbon when he wrote *The Decline and Fall of the Roman Empire*) elicited a powerful response. As Nobel prizewinner Saul Bellow's weary dean put it in *The Dean's December*: "It wasn't the barbarians, it wasn't the Christians, it wasn't moral corruption ... the real cause was the use of lead to prevent the souring of wine. Lead was the true source of the madness of the Caesars. Leaded wine brought the empire to ruin." But it was far more than a historical curiosity. "Now it's the whole world ... something Titanic, a Saturnian binge ... universal stupefaction, a gloomy murderousness, the raging of irritated nerves, human intelligence reduced by metal poison, so that the main ideas of mankind die out, including of course the idea of freedom."

Spokespeople for the tetraethyl lead, petrol and car industries — who for decades had kept public concern at bay by incanting Kehoe's idea of a balance — now found themselves engulfed by a tidal wave of outrage. Firmly on the side of the angels, Warren tells a triumphant story. What in the 1960s might have been envisaged as a programme for a century was accomplished by the turn of the millennium: American petrol became lead-free. To many, this was not enough. The only insurance against 'silent' lead poisoning was the universal screening of children for blood lead. Warren points to several political and economic developments which frustrated this advance. They are the justification for

his attention-grabbing opening statement, but here he is less convincing. Screening for everything is becoming an industry almost as powerful as that of motor cars.

It is the besetting sin of book reviewers to lament that the book reviewed is not the book they would have liked to write themselves. In this vein, Warren's book is what it says: a social history. It treats the vast and fascinating subject of clinical plumbism only briefly and barely touches on its far-ranging pathology, biochemistry, toxicology and treatment. Nor does it venture much beyond the United States. It remains a good story, nevertheless. ■

Thomas Dormandy is at 16 St Albans Road, London NW5 1RD, UK.

Metamorphosis of a brain

The Dying of Enoch Wallace: Life, Death, and the Changing Brain

by Ira B. Black

McGraw-Hill: 2000. 256 pp. \$24.95

Yadin Dudai

Like Kafka's Gregor Samsa, Enoch Wallace woke up one morning to discover that he had changed. But whereas Samsa's metamorphosis was instantaneous, the metamorphosis of Wallace was slow, sometimes insidious, yet in no way less frightening. This book is about the transformation of a healthy human brain into a demented one. Ira Black, a clinical neurologist and prominent neuroscientist, invites us to accompany Wallace on his ordeal. Black uses a proven literary device: one story as a frame around another. The frame-story tells us how a powerful, self-confident, 62-year-old New York investment banker turns within a few years into a helpless resident of an old people's care home. The core story is about the biological processes that might have been contributing to the dying of Wallace's brain.

The Dying of Enoch Wallace should be considered in the context of recent developments in the neurosciences. They all relate to the question of how the brain generates and sustains our humanity and individuality. Consider, for example, the following issue: we change dramatically throughout life, yet normally maintain a sense of persistence of self over time. This leads to the intuition that, in our unstable world, at least our brain of today is the same as our brain of yesterday and of tomorrow. Alas, science has deprived us even of this meagre consolation: in recent years it has become evident that Heraclitus' maxim *panta rei* — "everything flows" — applies to the brain as well. Not only do nerve cells and synapses change their shape and function, some of them are born or die

even while you are reading this passage.

The classical problem of 'the ship of Theseus' becomes inevitable: the mythical Greek hero's ship was placed on display in Athens, and over time parts of it were replaced, one by one, until none of the original remained. Is this still the same ship? Connections in our brain are continuously replaced — is this still the same brain? Well, it appears the same to us; science is challenged to demonstrate how this happens.

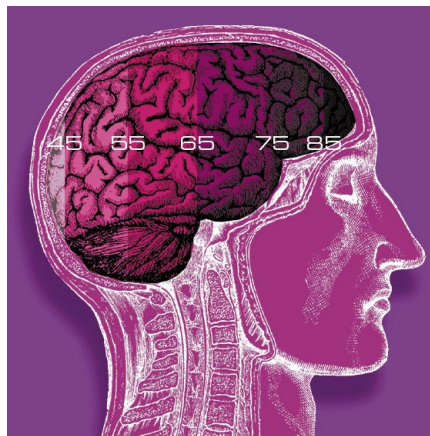
Even if we push metaphysics aside, the everlasting flux of brain tissue still leaves us with more tangible questions. How are our experiences encoded over time in an unstable substrate? How is an appropriate balance kept between birth, maintenance and death of connections in the brain? One of the conclusions of modern neuroscience is that development, response to stress, and learning and memory all share molecular and cellular universals. We thus become familiar with the molecular building blocks of the brain. We even get a notion of the algorithms executed by the neurons, and the circuits that are based on these building blocks.

But in most cases, we still miss the computational theories beyond it all. They should explain how stability is maintained in spite of the biological flux. In the meantime, the changing brain appears frail and ephemeral to us. And fragility is not merely a metaphor. We now believe that brains become demented when the balance between renewal, maintenance and death of brain tissue is disrupted. We certainly wish to know how this malfunction starts, and how we can stop it once it has been triggered.

The tragedy of dementia is the theme of the frame-story in this book. This will surely attract readers, because dementia is the terror of ageing societies. Only a century ago, less than 1% of the world's population was over 65 years of age. Now, it is already about 7%, and the prediction for 2050 is 15–20%. Therefore, ailments of old age such as dementias are becoming an epidemic.

The main concern is Alzheimer's disease. Named after the physician who first described it in 1906, it typically starts with recurrent lapses of memory, progressing to severe global memory deficits accompanied by other cognitive and emotional disturbances, and culminating in dissolution of personality and inability to perform even elementary bodily functions. This is what happens to the probably fictitious Enoch Wallace.

Black's account of this human misfortune is stereotypic and dry. We are left, hence, with the core story, which focuses on the discovery of growth factors in the nervous system, and their role in the development, maintenance and death of the brain. The description of how nerve growth factor and other neurotrophins were identified is interesting. The journey leads us to laboratories all



around the globe, where we get to meet some leading investigators.

The factual depiction of the history of this field is clear and straightforward. Black does attempt to colour the facts with information about the people involved, but unfortunately the fate of these accounts is similar to that of Wallace and his family in the frame-story. The attributes of the heroes include physical appearance, sometimes their hobbies, rarely conceptual incentives. For example, we are told (twice) that a certain leader in the field is a "lean, sinewy 6 feet 5 inches"; his colleague is a "willowy 6 feet, 2 inches"; another is "short and spare of build". Is neuroscience in the National Basketball Association? Some researchers are "wizards", "stars" and "geniuses", yet others are identified only by name; is this the evaluation sheet of the discipline?

The jury is still out on the causes of Alzheimer's. Possibilities include defective growth factors; faulty cholinergic neurons (nerve cells that secrete the neurotransmitter acetylcholine); and lesions in inter- and intracellular cascades, involving molecules such as amyloid precursor proteins and their so-called β -amyloid fragments, and resulting in the senile plaques and intracellular tangles that are the hallmark of Alzheimer's pathology.

Some important findings in the neurogenetics of dementia are missing from the book. This might be attributed to the rapid tempo of the field. It does, however, raise the question of when one should attempt to write the history of a discipline. It might also be due to the author's intention of focusing mainly on growth factors. However, the frame-story primes us to expect a broader account of what the dementias might be.

The seam between the frame and the core story is sewn in coarse, loose stitches. The text of the core story contains many technical terms and acronyms (all explained), and is accompanied by ample references, but this may not necessarily be attractive to the general reader. In spite of queries such as "Who are we? Where did we come from?", it is not a philosophical book. Despite the potential Kafkaesque connotation at the introduction,

literary fiction it is not. The glimpses of inside information about the scientific scene and the scientists who did the work do not suffice to ensure *Enoch Wallace* a prominent place on the shelf of scientific reportage: it lacks the thematic flow and captivating style of some other books in this expanding literary genre.

It is also not a personal account of how science is done; we miss an opportunity to really learn about the scientific culture at work from the point of view of a seasoned practitioner. The epilogue, expected to encapsulate the author's world view, occasionally reads too much like a proposal to a prospective donor: "how do we image a thought, an emotion, attention?" And the answer follows immediately: "Absurd religious questions [sic]. But neuroscience is providing the answer." Still, *The Dying of Enoch Wallace* is a story about what is going on in a prominent subdiscipline of brain research in which matters of development, learning, ageing, life and death intermingle. As such, it might provide an incentive to read more about these matters. ■

Yadin Dudai is in the Department of Neurobiology and Institute of Brain Research, Weizmann Institute of Science, Rehovot 76100, Israel.

Unravelling a tangled mind

Decoding Darkness: The Search for the Genetic Causes of Alzheimer's Disease

by Rudolph E. Tanzi & Ann B. Parson
Perseus: 2000. 320 pp. \$26, £15.95

Bruce A. Yankner

Few human conditions cause the kind of suffering associated with Alzheimer's disease, because it robs people of their humanity. Despite substantial progress in Alzheimer's research over the past 15 years, the cause of the disease has yet to be definitively established, and there is still little that can be done to halt the inexorable march of cognitive decline. *Decoding Darkness* provides an autobiographical perspective on the discovery of the genes that cause familial Alzheimer's disease, and the evolution of the major pathogenetic hypotheses, by one of the leading researchers in the field.

A substantial portion of the book is devoted to the experiences of Rudolph Tanzi, a scientist who has made important contributions to the genetics of Alzheimer's disease. These personal adventures add entertainment value to the story. The book's coverage ranges from details of the positional cloning of genes that are associated with Alzheimer's disease to the Calistoga mud baths. It is written in a conversational style, with a

minimum of technical jargon, and is an easy read even for non-scientists.

Although the book is highly personal, journalist Ann Parson made extensive efforts to interview many scientists in the field repeatedly, to get their stories right. The end result is a book that portrays many different viewpoints. Although the underlying theme is that amyloid protein is probably the cause of the disease, most of the major areas of investigation and points of view are represented.

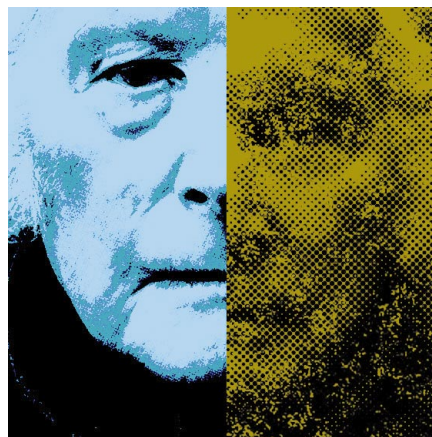
Tanzi and Parson begin by describing some of the first attempts at the positional cloning of a disease gene in James Gusella's lab at the Massachusetts General Hospital in the early 1980s. The technique of linking benign genetic variations with a disease locus was a major advance in genetics that initially led to the mapping of genes for Huntington's disease and Duchenne muscular dystrophy. These initial forays gave rise to the search for genes that might cause Alzheimer's disease.

The modern era of Alzheimer's research began with George Glenner's discovery in 1984 that the protein deposits in the brains of patients with the disease are composed of amyloid- β protein. Glenner demonstrated that this same peptide accumulated in the brains of individuals with Down's syndrome, many of whom became demented as they reached late middle age.

With subsequent cloning of the gene for the amyloid precursor protein (APP) and its localization to chromosome 21, the Down's syndrome connection acquired new significance, because it was suspected that mutations in this gene might be the cause of familial cases of Alzheimer's. This idea was supported by a report in the late 1980s demonstrating genetic linkage of several Alzheimer's disease families to chromosome 21. Ironically, this paper turned out to be incorrect; the genetic defect in these families proved to be linked to chromosome 14, and the causative genetic mutations were in the presenilin genes rather than in the genes encoding APP.

Tanzi and Parson describe how a period of uncertainty followed these initial studies, while investigators looked elsewhere for the genes and proteins behind Alzheimer's disease. Enthusiasm for amyloid- β protein as a causal element in the disease was revived in the early 1990s with the discovery that the protein has neurotoxic properties. In 1992, John Hardy and colleagues described the first APP mutation in Alzheimer's disease families, and the description of other APP mutations soon followed. These findings provided indisputable evidence that the disease could start with a defect in APP. The power of genetics to provide instant proof of the relevance of a protein to the disease became all too clear.

But the mechanism by which genetic mutations result in Alzheimer's disease was



not as clear as for some of the other inherited degenerative conditions, such as Huntington's disease and Duchenne muscular dystrophy. An important insight came from the observation by Steven Younkin and colleagues that APP mutations resulted in a subtle, yet important change in the production of amyloid- β protein. The resulting protein was longer than normal and had a high propensity to aggregate and form plaques in the brain. This peptide was presumed to be more neurotoxic and a key factor in the disease process. In 1995, scientists at Athena Neurosciences and Exemplar Corporation showed that expression of an APP mutation in transgenic mice resulted in amyloid plaque formation in the brain. This model solidified the connection between the molecular genetics and the pathology of the disease, and provided an animal model that could be used for drug testing.

A new gene family associated with early-onset Alzheimer's disease, the presenilins, was discovered by Peter St George-Hyslop and colleagues, signalling a second major chapter in the neurogenetics of Alzheimer's disease. The book gives a detailed account of the intense race between several groups to find the gene. Tanzi's own group had been working on chromosome 14, which contains the presenilin-1 gene, and had come very close.

Disease-causing presenilin mutations were found to increase the production of the longer, more pathogenic form of amyloid, similar to the effects of disease-causing mutations in APP. In addition, the presenilins turned out to be associated with interesting biology, playing a pivotal role in brain development and potentially affecting signalling pathways involving cell death. Furthermore, recent findings raise the intriguing possibility that presenilin may be one of the two enzymes that mediate the generation of amyloid- β protein. Thus, altered amyloid metabolism is currently the most straightforward explanation for the pathogenic effects of presenilin mutations, although whether this is the sole or predominant mechanism is not known.

Despite their conceptual importance, mutations in the presenilins and APP account for a relatively small proportion of cases of Alzheimer's disease. The major known genetic risk factor is the $\epsilon 4$ variant of the plasma protein apolipoprotein E (apoE), discovered by Allen Roses and colleagues. Unlike the mutations in APP and the presenilins, which almost guarantee that a person will develop Alzheimer's disease, the apoE $\epsilon 4$ variant is a susceptibility gene that increases the risk of Alzheimer's disease after the age of 60. But the mechanism by which apoE causes Alzheimer's is not known. The discovery of the role of apoE was followed by reports of several other potential susceptibility genes, one of the most hotly debated being α_2 -macroglobulin, a protein that may be involved in the metabolism of amyloid- β protein.

The convergence of genetics and cell biology gave rise to the amyloid hypothesis, which posits that the abnormal generation, localization or aggregation of this peptide is toxic to neurons and leads to dementia. However, many believe that other factors, such as disruption of the cell infrastructure by neurofibrillary tangles, are more likely to be directly related to dementia. This idea has recently received support from the identification of mutations in the tau protein — the major component of neurofibrillary tangles — in a group of dementing diseases known as frontotemporal dementias. Genetic abnormalities in tau can therefore lead to neuronal degeneration, raising the possibility that abnormalities of tau in Alzheimer's disease may have similar effects.

If the amyloid hypothesis is correct, we may be on the verge of finding drugs that alter the course of the disease. The enzyme responsible for one of the two cleavages that generate amyloid- β from APP has recently been identified, making it a matter of time before effective inhibitors of the process become available.

Another novel approach has been to use immunization with the amyloid peptide. This dramatically reduces subsequent plaque formation in mouse models of the disease. And the amyloid vaccine has now entered phase I clinical trials. Additionally, therapeutic agents designed to prevent amyloid aggregation and toxicity are being developed. The hope is that these approaches will result in drugs that affect the cause of the disease rather than just the symptoms, a limitation of currently available drugs.

Decoding Darkness will appeal to those who want to know about the human dimension behind this exciting science. It is recommended reading for anyone interested in Alzheimer's disease and neurogenetics. ■

Bruce A. Yankner is in the Department of Neurology, Harvard Medical School, and the Division of Neuroscience, Children's Hospital, Enders 260, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

What's in a name?

Linnaeus' marginal jottings created order out of botanical chaos.

Sandra Knapp

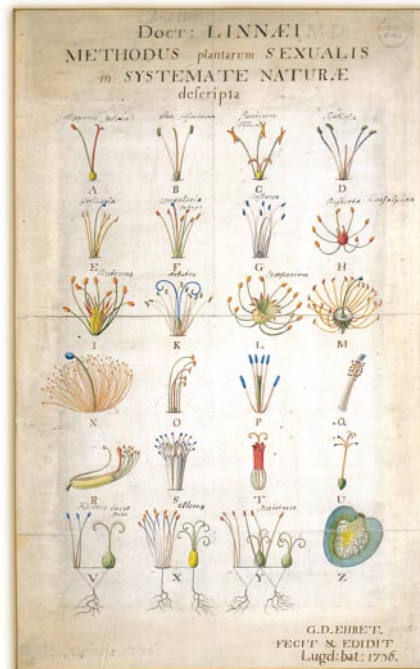
Lamenting her love for the 'wrong' man, Juliet muses, "What's in a name? that which we call a rose By any other name would smell as sweet". She asserts that Romeo's 'essence', his goodness, desirability and character, transcends his name. An Elizabethan botanist in the audience at the Globe theatre might have agreed, not because he (and he surely would have been a man) believed names were of no consequence, but because he too would have relied on defining the essence of things.

Humans have invented many different systems of naming. Throughout the Dark Ages in Europe, monks copied and re-copied the treatises of the Greek natural philosophers, keeping their naming systems alive. Aristotelian logic, a force in the natural sciences until the mid-nineteenth century, required that names reflect hidden reality or essence. A name thus consisted of a generic part (*character naturalis*), the essence of the organism, and a specific part (*differentia*), distinguishing it from all others in the genus. Names were usually Latin phrases, polynomials, but could be a single word. Consistency was not the order of the day.

Revolution arrived in the eighteenth century in the shape of Carl Linnaeus, who has been called many things — the first ecologist, a botanical pornographer and a "compleat naturalist". It was said, "God created; Linnaeus arranged", and the man himself was certainly not reticent about his own abilities. The task he undertook was to catalogue all life. Being a medical doctor and a botanist, he naturally concentrated on plants.

As Europeans began to explore the rest of the world, they brought back organisms from far-away lands — especially plants, which could be grown easily in the gardens and hothouses of rich patrons of science. The sheer diversity may not be so remarkable to us today — Linnaeus' catalogue of the world's flora comes to only about 9,000 taxa — but it severely strained the polynomial system of naming. To differentiate one or two similar things is easy, but when the numbers increase, names become cumbersome and impossible to memorize.

Linnaeus, for whom brevity was an ideal — "for beauty and perfection science requires conciseness and brevity" — invented a system of *nomina trivialia*, trivial names, to run in parallel with the phrase-names required by Aristotelian logic. His greatest botanical work, *Species Plantarum*, was published in 1753, after 20 years of labour. It was



Botanical pornography? Georg Ehret's drawing of the Linnaean 'sexual system' of plants.

a complete catalogue of all known plants but in it Linnaeus used for the first time his trivial names as marginal notes — not to replace 'proper' names, but as a way of allowing botanists to mentally organize the ever-increasing diversity. Thus, the tomato had a botanical name of "*SOLANUM caule inerme herbaceo, foliis pinnatis incis, racemis simplicibus*" (*SOLANUM* with a smooth herbaceous stem, incised pinnate leaves and a simple inflorescence), which was equated with earlier botanists' names — Bauhin's 1620 "*Solanum pomiferum, fructu rotundo striato molli*" (apple-like solanum with a round, smooth, striped fruit) and Mattioli's 1586 "*Poma amoris*" (apple of love). For convenience, in the margin of the text, Linnaeus gave the tomato, in the genus *SOLANUM*, the trivial name of "lycopersicum". Despite an initially cold reception from the learned, the simplicity of Linnaeus' system recom-

By linking a name with a single specimen, biology becomes a repeatable science.

mended it to botanists and gardeners alike, and by the time he died in 1778 binary nomenclature was firmly established.

Nomenclature is often dismissed as the arcane, sterile province of academic taxonomists. But without a name, we find it impossible to communicate about even common objects — witness the confusion caused by an English child asking for a rubber in a US classroom. Biological nomenclature as we know it today began with Linnaeus. Our ability to call a rose, not by any name, but by a precise name such as *Rosa canina* or *Rosa multiflora*, allows us to communicate in a common language. By referring to specimens or drawings, Linnaeus tried to link the name with a real object. The type method, devised as a way to control the application of names in the early part of the twentieth century, further refined this link. By linking a name with a single specimen, which can be examined again and again, biology becomes a repeatable science. Scientific names, the genus and species, provide reference points in biological space to which we may compare future discoveries and information.

But although names imply an underlying order, it is folly to equate them with 'biological reality'. Our concept of how organisms are related differs radically from that of Linnaeus or that of Charles Darwin, and is improving all the time. Concepts of what a species actually is are also changing as knowledge increases — not only knowledge as to the extent of diversity, but also that of the genetic structure inherent in all living things. So names change, but rather than being irritated that the tomato is now known as *Solanum lycopersicum* rather than *Lycopersicon esculentum* (interesting that Linnaeus got it right, at least according to today's experts), we should reflect on the fact that an advance in knowledge underpins the change. Nomenclature comprises more than just species and genus names, however, and debate rages as to how biologists should reflect hierarchy and classification in their naming systems.

By inventing a simple, concise method of naming organisms by genus and species, Linnaeus revolutionized biology. How much more useful to know that the creature from whom we obtain DNA sequence is *Drosophila melanogaster* — comparable to other such flies, its identity verifiable and our results repeatable — rather than one of Jorge Luis Borges's creatures "that from a long way off look like flies".

Sandra Knapp is in the Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

The stars my incarnation

You read it here first, folks.

Robert A. Metzger

“Welcome back, Creator Metzger.” I attempted a groan, but the noise sounded like a cat hacking up a fur ball. Before me stood several dozen black-robed priests, the nearest one dropping to his knees and shuffling toward my optical receiver.

“On this 200th anniversary of our arrival at the Centauri system, we wish to tell you of our progress, of our further implementation of Your Vision.”

I angled my optical interface upward, peering up through the transparent ceiling of the Cathedral. Night, and Alpha Centauri A burned bright, less than one light day away.

“Get over here,” I said to the priest.

He knee-walked to the optical interface.

“I’m not your Creator,” I said. “I was a little-known science-fiction writer of the early twenty-first century who had the great misfortune of getting published in a journal called *Nature*.”

The priest hid his face in his hands, and shuddered. “You gave us the stars, showed us where our destiny lay.”

I sighed, this time the speaker emitting a bark that sounded like a car backfiring. “Every time one of you wakes me up, I try to explain this to you, try to make it clear, try to set the record straight, but none of you seems to get it.”

“We are not worthy,” he said as he dropped to the crystal floor, grinding his nose into its rainbow-coloured surface.

“I’ll try once again,” I said, doubtful that

this time would be any different. “It was the year 2000, nearly 700 years ago, and *Nature* was celebrating the beginning of the new millennium with essays about what the next 1,000 years might bring. I wrote one of those pieces, about a future in which most of the Solar System had been transported four light years away to the Centauri star system.”

“And Your Will was done!” screamed the head priest.

“Yes, my will was done. In that fictional piece, I proposed that if the Sun’s magnetic fields could be controlled, then they could be used to expel a relativistic jet of protons, basically turning the Sun into a rocket engine. Under an extremely small but steady acceleration, just one ten-thousandth of a standard gee, after 200 years the Sun would be moving at nearly 2% the speed of light and would have travelled two light years. At that point, the direction of the acceleration would be reversed, and 200 years later the Sun would come to a halt in the Centauri system. The whole trip would consume only about 1% of the Sun’s mass.

“Of course it wouldn’t be an easy trip. Even under that gentle acceleration the big outer planets would be left behind, and Earth’s circular orbit would be turned into an ellipse. As a result, half of the year the planet would freeze and the other half it would bake. It seemed to me that the easiest way to get the job done would be if the Sun first achieved self-awareness.”

“Your Vision was made real.”

“Yes. I died back in 2027, the result of a Man-versus-Speeding-Bus event. They iced me down, keeping me in cold storage for

spare parts. The next 50 years saw the development of self-awareness in silicon-based hardware, which then gave way to self-awareness in plasma systems, the density of ions twirling about in a magnetic field far in excess of anything silicon was capable of producing.

“In 2083, one totally certifiable loon named Rufus Mapelton, a plasma hacker who spent every conscious and unconscious moment interfaced to the plasma network, stumbled across my long-forgotten *Nature* piece. Mapelton had a little ‘eureka’ moment, realizing that if he could transfer a self-aware plasma-based entity into the photosphere of the Sun, it could then take control of the Sun’s magnetic fields and implement my fictional vision.”

“He was Your Instrument.”

“I suppose so. Mapelton synthesized his plasma virus, uplinked it to one of the Sol-Corona satellites, and then had the satellite spiral into the Sun. Six weeks later the Sun achieved self-awareness and decided to head off for the Centauri system. It was then that they reconstructed what was left of my mind, hoping that I could help shut down the Sun’s desire for a new home. Of course, I couldn’t. Between the quakes, the comet impacts and the skewed orbit, intelligent life barely survived the trip.”

“And now we flourish,” said the priest. “And we wish to tell You of our further achievements.”

It was always like this. They listened, but never really heard me. “So what’s new?”

“We plan a wondrous journey.”

“Oh?”

“What few organics remain have decided to transfer their consciousness into the Sun’s photospheric magnetic matrix. Once the transfer is complete, we will resume the Sun’s journey, moving toward the Galactic Centre, transferring a portion of ourselves into each star we pass.”

“I’d also mentioned that in the *Nature* piece, hadn’t I?” I asked.

The priest nodded.

And I knew what would finally come. Each and every star would eventually achieve self-awareness, each setting off on its own journey of discovery and exploration. And in the end, the Universe itself would be aware, the organic having given way to the plasma, the Universe itself forever transformed by the ramblings of an insignificant twenty-first-century science-fiction writer. ■

Robert A. Metzger is a science-fiction writer. He has just completed a novel, *Picoversion*, and looks forward to the fame awaiting him in the twenty-eighth century.



JACEY

Small but mighty timekeepers

André Adoutte

A tiny RNA molecule ensures that the larvae of a roundworm develop into adults. The discovery of this RNA in many other animal groups implies that this way of keeping developmental time may be universal.

How is the timing of major developmental transitions in animals genetically controlled? How, for instance, does a larva know when to transform into an adult? One possible answer was suggested by the discovery of 'heterochronic' mutations in a small, intensely studied roundworm, the nematode *Caenorhabditis elegans*. These mutations, in a series of genes, either advance or retard the schedule of many developmental events as the worm progresses through its larval stages¹.

Some of these 'timekeeping' genes — such as *lin-4*, which acts between larval stages 1 and 2, and *let-7*, which is involved in the transition from larval stage 4 to adult — work in a rather quirky way. Most genes encode long messenger RNAs, which are translated into protein (Fig. 1a). But genes like *lin-4* and *let-7* instead encode small RNAs that inhibit the expression of a quite different gene, probably by stopping the translation of the mRNA it encodes (Fig. 1b). It is the inhibition of the target genes that sets the genetic cascade controlling a particular developmental transition. These genes, by acting on yet another set of genes, are the major regulators of the transition in question^{2,3}. But ultimately, the precise timing of these events is achieved by regulating the amount, and the time of synthesis, of the small RNAs^{2,3}.

From bacteria to mammals, there are precedents to the use of RNA for controlling gene expression⁴. Unlike most groups of animals, however, nematodes develop through a strictly defined pattern of cell divisions to yield precise cell lineages. So, as neat as this particular mechanism seems, it might simply have been a developmental timing device specific to nematodes.

But on page 86 of this issue⁵, Pasquinelli and colleagues show that one of the nematode small RNAs, that encoded by the *let-7* gene, is amazingly conserved across all the major groups of bilaterally symmetrical animals — from flies to vertebrates, and from worms and molluscs to echinoderm larvae. Moreover, *let-7* is expressed during the development of these different animals at times — such as just before metamorphosis in the fruitfly *Drosophila melanogaster* — that suggest it is involved in controlling a major developmental transition. Finally, because the genomes of some of the animals tested have been completely sequenced, or

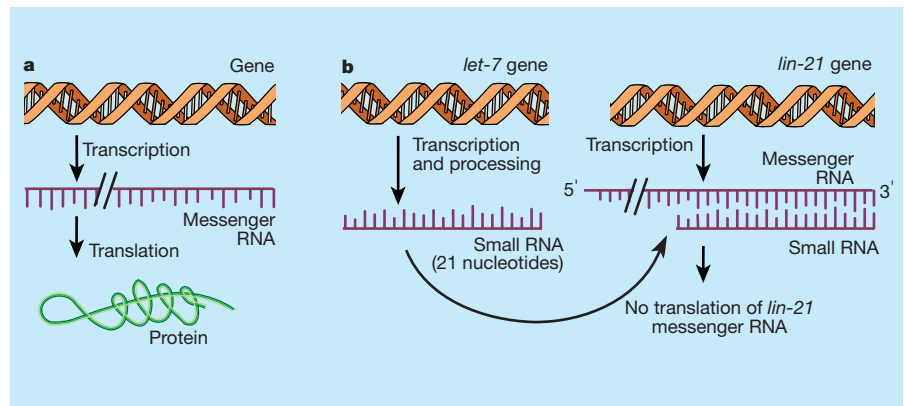


Figure 1 Keeping time with small RNAs. **a**, Conventionally, genes are transcribed into messenger RNAs, which are then translated into proteins. It is the proteins that do the 'work' specified by the gene sequences. **b**, A few genes, such as the nematode *let-7* gene (left), are instead transcribed in response to upstream signals (not shown) and processed to produce small RNAs — the *let-7* small RNA is a mere 21 nucleotides in length. Here it is the small RNA that does the work: it pairs up with the untranslated nucleotide sequence at the 3' untranslated end of a target messenger RNA (the *lin-41* RNA in the case of *let-7*), and probably prevents it from being translated into protein. This relieves the inhibition on other genes, setting off a genetic cascade (not shown) that steers the organism through a major developmental transition. This mechanism of gene inhibition by small RNAs is probably not the same as 'RNA silencing', which also involves small RNAs but results in the degradation of the target RNAs. Pasquinelli *et al.*⁵ have found the *let-7* small RNA in all major groups of bilaterally symmetrical animals, a result that hints that this mechanism of developmental timekeeping is also conserved.

almost so, the authors were able to identify *let-7*'s target gene in both *Drosophila* and vertebrates. This result hints that the whole genetic cascade has been conserved during evolution.

The degree of conservation of the *let-7* small RNA is quite remarkable⁵. In its mature form, the *C. elegans let-7* RNA comprises just 21 nucleotides. *Drosophila* has one exact match, and humans have three, as well as another small RNA in which 20 of the 21 nucleotides are identical. This renders these small RNAs among the most highly evolutionarily conserved molecules — similar levels of conservation are found only in stretches of ribosomal RNA or small nucleolar RNAs. The degree of conservation hints that there are considerable constraints on the nucleotide sequence of the *let-7* small RNA. One such constraint could be the fact that other RNAs (such as the *lin-41* mRNA in *Drosophila*, vertebrates⁵ and *C. elegans*) and proteins interact with *let-7*. This in a sense 'freezes' the evolution of the RNA-protein complex, because any change in one of the partners would require compensatory changes in the other. Indeed, the target

sequence on the interacting mRNA is also highly conserved. The conservation of the size of *let-7* is in itself not too surprising, however. This RNA is as small as it could be — it has barely enough nucleotides to establish a thermodynamically stable pairing with its target mRNA.

The remarkable conservation of this timing device raises a host of questions. For example, how might a system involving such a tight interaction between two complementary RNA sequences have come into existence? This 'invention' probably occurred in the last common ancestor of bilaterally symmetrical animals, as the other descendants of this ancestor — sponges and cnidarians (such as jellyfish) — are apparently without it. It probably started as a gene duplication, which somehow became subverted for use in controlling translation.

Another mystery is how the binding of a small RNA at the untranslated end of the target mRNA affects that mRNA's translation. Also, increased expression of the *let-7* gene sets off the genetic cascade, but how is *let-7* itself regulated? And which are the 'effector' genes that actually produce the developmen-

tal transition? But perhaps the most intriguing question is whether the whole regulatory cascade — including the effector genes — is the same across animal groups.

In the late 1980s, biologists were baffled by the discovery of the extraordinary conservation of 'developmental genes' across very distant animal groups. Organisms as distantly related as *Drosophila* (an arthropod) and mice and humans (vertebrates) were found to share a gene array known as the Hox complex. One of the functions of this complex is to specify the destiny of different parts of the body along the head-to-tail axis of these animals. Since then, the list of similar gene networks that are conserved across animal phylogeny has been steadily increasing. The implication is that the genes were already present in the last common ancestor of all the animals that share them today — it is unlikely that identical genes would be 'invented' independently in different lineages.

But the key question is whether the processes in which these genes are involved are put to the same use in different animals. This would indicate that the processes were also present in the last common ancestor. Alternatively, the genes might have been 'co-opted' for use in quite different functions in the varying animal lineages. The same question applies to the *let-7* small RNA and its genetic cascade. With time, the regulators and effectors of *let-7* in several phyla will no doubt be identified. Then we will be able to answer the question of whether this time-keeping mechanism is truly universal. ■

André Adoutte is at the Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette, France.
e-mail: adoutte@cgm.cnrs-gif.fr

1. Ambros, V. *Curr. Opin. Genet. Dev.* **10**, 428–433 (2000).
2. Reinhart, B. J. *et al. Nature* **403**, 901–906 (2000).
3. Slack, F. J. *et al. Mol. Cell* **5**, 659–669 (2000).
4. Eddy, S. R. *Curr. Opin. Genet. Dev.* **9**, 695–699 (1999).
5. Pasquinelli, A. E. *et al. Nature* **408**, 86–89 (2000).

Solid-state physics

Surprising movements in solids

Ulrich Gösele

A process as simple as diffusion should be easy to understand. But our knowledge of the movements of atoms in semiconductors is still far from complete.

The natural movements (diffusion) of atoms in a solid are made possible by defects in the crystal lattice. So studies of diffusion in technologically important materials, such as semiconductors, help

engineers to make smaller and better microelectronic and optoelectronic devices. On page 69 of this issue Bracht *et al.* report that in the semiconductor compound gallium antimonide the antimony atoms move a thousand times slower than the gallium atoms. This result contradicts earlier measurements, and was not expected from the diffusion behaviour of other semiconductor compounds.

Gallium antimonide is used to make infrared lasers and photodetectors, night-vision devices, solar cells and some special transistors. These are niche applications compared to those of the ubiquitous silicon, or gallium arsenide and indium phosphide, which are mostly used in lasers and high-speed telecommunications devices. The significance of Bracht and colleagues' result does not stem from the technological importance of gallium antimonide. Rather, it is because the finding is another in a long line of unexpected differences and similarities encountered in the history of understanding diffusion in semiconductors.

It was first suggested by the Russian scientist Frenkel² in 1924 that the movement of atoms in a crystalline material may not just occur through two neighbouring atoms exchanging positions, but that defects are required. In metals, these defects were shown to be missing lattice atoms, known as vacan-

cies. How fast a lattice atom can move depends on the concentration of these vacancies — which, for thermodynamic reasons, are present in increasing concentrations with increasing temperature — and on the mobility of the vacancies.

When semiconductors came along in the 1940s, germanium was used to make the first transistors. Germanium atoms were also found to diffuse with the help of germanium vacancies, as happens with metals. Technological interest soon shifted to crystalline silicon, which has several advantages over germanium for making electronic devices: for example, it can form a high-quality electrically insulating oxide, and it can be grown as large single crystals (Fig. 1).

Unexpectedly, the mechanism of diffusion of silicon atoms in silicon crystals — also known as 'self-diffusion' — turned out to be different from that in metals and germanium. As well as the usual vacancy mechanism, extra silicon atoms called self-interstitials play a prominent role³. But even though Frenkel had already discussed self-interstitials in his early papers, the idea of vacancies dominating diffusion processes was so engrained that it took more than 20 years for the role of self-interstitials to be accepted and finally incorporated into computer models used to design microelectronic devices.

Experiments using silicon later showed that the coefficients of self-diffusion (the rates of self-diffusion) for both of these defect mechanisms are extremely similar in magnitude over an extended temperature range of several hundred degrees Celsius. This would have been astonishing enough, but we also learned that the individual factors determining self-diffusion in silicon — the defect concentrations at a given temperature, and the mobility of these defects — are almost unbelievably similar. For example, at the melting point of silicon the concentrations of vacancies and self-interstitials differ by only about 30%. This surprising fact is at the heart of the observation that, during growth of silicon crystals, either vacancies or self-interstitials can agglomerate, and so both are responsible for forming undesirable defects⁴.

It also offers the possibility of growing the crystals in such a way that almost all the vacancies and self-interstitials annihilate each other, so that practically no defects remain. This feat of engineering is key to producing the almost perfect silicon wafers required for the next generation of computer chips. From a scientific point of view we still do not know whether these unexpected similarities in silicon are purely accidental or whether some deeper scientific truth lies behind them.

The situation is more complex for semiconductor compounds, however, because there are two sublattices — in gallium arsenide, one for gallium atoms and one for

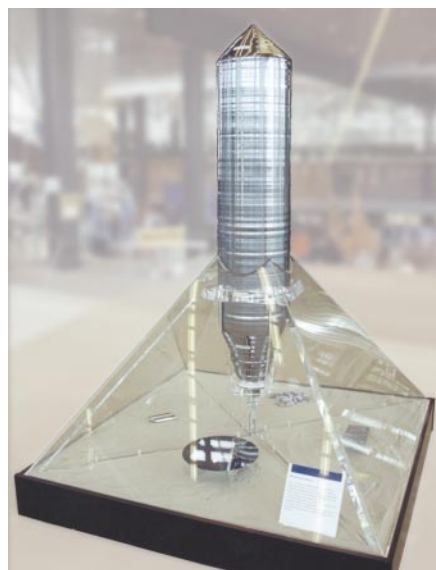


Figure 1 One of the largest available single crystals of silicon, presented at an exhibition of the German Physical Society. It symbolizes a modern-day version of the 'Stone of Wisdom', the topic and theme of the exhibition. (Photo courtesy of Research Center Jülich.)

arsenic. How do gallium and arsenic atoms move in these sublattices? Fortunately, what we have learned in the case of silicon can also be applied to semiconductor compounds, in an appropriately modified form. There are vacancies and self-interstitials in both sublattices, and both types of defect play a role in diffusion processes. One of the first unexpected observations was that the movement of atoms on both sublattices is generally similar in magnitude. One possible reason for this is that one atom on one sublattice (say, gallium) is typically surrounded by four atoms of the other sublattice (arsenic) and vice versa. Although the binding between the atoms is not symmetric, the whole diffusion process seems to involve some kind of averaging over the different configurations and local atomic environments, ultimately favouring similar diffusivities on both sublattices.

Earlier self-diffusion measurements for gallium and antimony in gallium antimonide⁵ found the diffusivities of gallium and antimony to be similar, as expected. But Bracht *et al.*¹ now report experiments showing that antimony diffuses 1,000 times slower than gallium. Immediately two questions arise. First, are the new results more reliable than the earlier ones? And second, supposing the new results are true, what makes this material different from other semiconductors?

The first question can be answered easily. Previous measurements were based on radioactive elements diffusing in from the

surface, whereas the technique of Bracht *et al.* involves growing semiconductor structures containing specific non-radioactive isotopes. The existence of these isotopes throughout the structure avoids the undesirable surface effects associated with radioactive probes, and leads to more accurate measurements. Because most of the self-diffusion data from other semiconductor compounds are also based on radioactive tracers, perhaps we should also worry about the reliability of these other established data.

Concerning the second question, Bracht *et al.* speculate that some of the gallium atoms move into empty antimony sites, thereby blocking the vacancy-assisted movement of antimony atoms, and at the same time increasing the concentration of gallium vacancies. This process would require unusual interactions of native defects. The model proposed by Bracht *et al.* should be investigated by quantum-mechanical calculations of these diffusion processes, and might inspire new thinking on diffusion mechanisms in semiconductors. This could lead to new insights into the behaviour of materials other than gallium antimonide. ■

Ulrich Gösele is at the Max Planck Institute of Microstructure Physics, Weinberg 2, D-06120 Halle, Germany.

e-mail: goesele@mpi-halle.de

1. Bracht, H. *et al.* *Nature* **408**, 69–72 (2000).
2. Frenkel, Y. Z. *Physik* **26**, 117–138 (1924).
3. Seeger, A. & Chik, C. P. *Phys. Stat. Sol.* **29**, 455–542 (1968).
4. Voronkov, V. V. J. *Cryst. Growth* **59**, 625–643 (1982).
5. Weiler, D. & Mehrer, H. *Phil. Mag. A* **49**, 309–325 (1984).

of its parent. For position-dependent events, the identity of a cell is controlled by its position within an organ. Typically, clonal events are common early in development, occurring in meristems — the clusters of undifferentiated cells that spawn the developing organs. Positional events occur later, and can involve the switching of cells between different fates. Both clonal and positional effects come into play in the endosperm, and Becraft and Asuncion-Crabb¹ flesh out the bones of our knowledge of the positional effects.

The endosperm, a major component of the cereal grain, originates from a fertilization event (of the maternal ‘central’ cell by a paternal sperm cell) that occurs in parallel to that which produces the embryo. The cereal endosperm consists of two basic cell types. The inner, energy-storage tissue is composed of starchy endosperm cells. These are surrounded by one to three layers of aleurone cells — protein- and oil-rich cells that secrete enzymes, allowing the mobilization of endosperm reserves during seed germination.

The first step in the formation of these cell types is the division of nuclei arranged around the periphery of the ‘endosperm sac’. Each nucleus gives rise to a starchy endosperm ‘initial’ and an aleurone initial² (Fig. 1a, b, overleaf), from which the two cell types originate after further divisions (Fig. 1c). These latter divisions could be seen simply as a clonal process. Alternatively, the further development of the aleurone initials might require a signal or signals from the surrounding maternal tissue — a positional effect.

Unusually, the endosperm is seen as an ‘end-cell’ system, in that it is not viable after the seed matures. Trying to persuade endosperm cells to divide further in tissue culture leads to a loss of the differentiated states of the two cell types. These observations suggest that endosperm cells are not plastic, that is, they cannot switch their fate. But Becraft and Asuncion-Crabb¹ now show that this is not the case.

Using a common technique for inducing chromosome breaks, the authors eliminated the function of a gene called *Dek1* in particular sectors in maize plants. In the resulting genetic mosaics, aleurone cells assumed the identity of starchy endosperm cells (Fig. 1c). The authors were then able to restore both *Dek1* function and aleurone-cell identity to peripheral starchy endosperm cells in the affected sectors. As well as showing that cell fate is not fixed until late in endosperm development, these results reveal that the *Dek1* gene is required to respond to positional cues that send early aleurone cells further down the road to specialization.

Becraft and Asuncion-Crabb went on to reveal further positional effects that fix the identity, and control the further differen-

Plant biology

Turning fields into grains

Richard D. Thompson

Cereal seeds contain an embryo and an endosperm, which is used as a food source. Differentiation of the endosperm is guided by several ‘positional’ molecular cues throughout development.

Which food supplies some 60% or more of the calories in the human diet? The answer is cereal plants — specifically, the endosperm of cereal seeds. The endosperm is a simple, nutrient-supplying tissue that surrounds the embryo of higher plants, and is composed of two types of cell: the protein- and oil-rich outer cells, and the starchy inner cells.

Writing in *Development*, Becraft and Asuncion-Crabb¹ shed light on the factors involved in the development of these cell types. It used to be assumed that the characteristics of the cells are fixed two to three days after pollination. But, by re-examining genetic mosaics in the endosperm, Becraft and Asuncion-Crabb show that the fate of these cells can be changed at least up until

the last cell division, as late as 22 days after pollination. The authors also identify three classes of ‘kernel mosaic mutant’, which correspond to three stages in the differentiation of the outer cells. The results might one day help in the manipulation of the oil, protein and carbohydrate content of cereal grains.

As plant cells are usually surrounded by rigid cell walls, there is little scope for them to migrate. So the shape and structure of plant organs is determined by the number of cell divisions and the extent of cell elongation. Similarly the specialization, or differentiation, of cells to perform different functions within a plant incorporates two types of event. For clonal events, the identity of a newly generated cell is determined by that

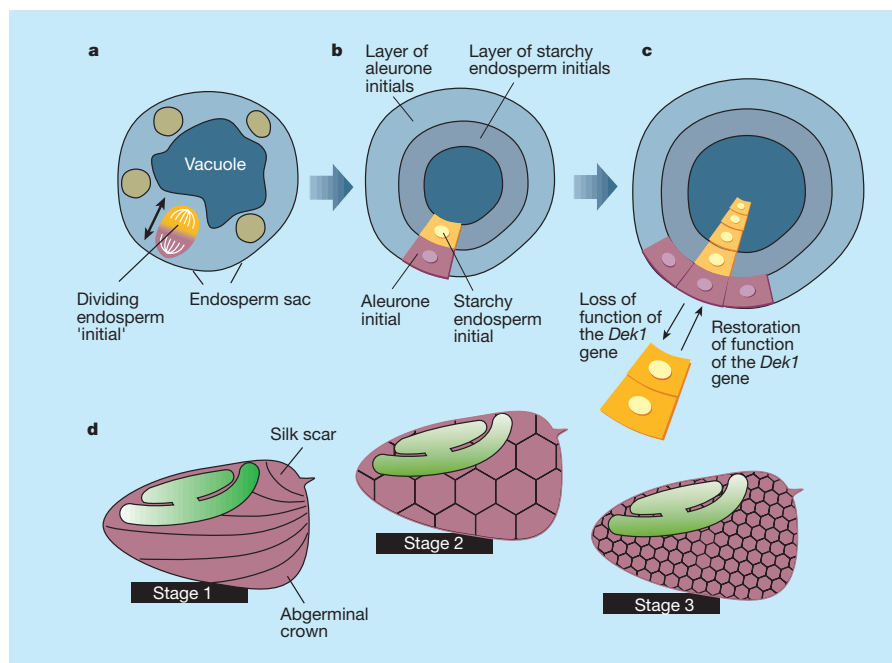


Figure 1 Deciding the fate of seed endosperm cells. **a, b**, Endosperm 'initial cells' (**a**) divide to give rise to aleurone and starchy endosperm initials (**b**; for simplicity, only one of each is shown). The initials then divide further, perpendicular to each other. **c**, The resulting cells differentiate to form aleurone and starchy endosperm cells. Becraft and Asuncion-Crabb¹ show that positional cues are required to specify aleurone cells. Loss of function of the *Dek1* gene leads to the conversion of aleurone cells to starchy endosperm cells. When gene function is restored, the outermost endosperm cells revert to aleurone cells. So the fate of an aleurone cell is not fixed; positional cues are required, and *Dek1* responds to these cues. **d**, Becraft and Asuncion-Crabb also identified three broad categories of mutant maize plant, in which different stages of aleurone-cell differentiation are affected. The authors propose a three-stage differentiation process. Stage 1, the identity of the aleurone cell is controlled by its position within a gradient — from the seed's 'silk scar' to its 'abgerminal crown' — of the factor(s) affected in the first category of mutant plant. Stage 2, a few large fields of cells form, and cell fate is determined by position within those fields. Stage 3, many smaller fields form. Purple, endosperm; green, embryo. **d** is modified with permission from ref. 1.

tion, of aleurone cells. The authors looked at a variety of maize 'kernel mosaic' mutants that had regions in which the development of aleurone cells was abnormal. They then determined the extent of differentiation in the mutant sectors, and suggest that each of the mutant genes acts in one of three successive developmental stages (Fig. 1d).

In stage 1, an aleurone cell's identity is controlled by its position in a gradient, of unknown composition, across the inner part of the seed; this gradient is disrupted in stage-1 mutants. In stage 2, the further differentiation of an aleurone cell is determined by its position in relatively few large fields throughout the endosperm. During stage 3, the fate-determining fields are smaller and more numerous. The origin and nature of these developmental fields is not yet known: they may reflect cell groups connected by plasmodesmata (threads of cytoplasm), or they might originate from limiting dilution of some component(s) from either the sperm cell or the central cell.

The layer of developing starchy endosperm cells contains several discrete domains — with different biochemical properties

and functions — that are characterized by their expression of particular 'marker' proteins^{3,4}. One domain, the 'basal transfer layer', corresponds to a region specialized to promote solute uptake into the seed. The function of the second cellular domain, the embryo-surrounding region, is not yet clear^{3,4}. We know little about whether or how these cell types can switch fates. But if they could be experimentally induced to change their destinies, the potential for changing the composition of the seed to suit the dietary needs of humans would be considerable. Likewise, one spin-off of Becraft and Asuncion-Crabb's¹ findings might be a better ability to alter the proportion of protein- and oil-rich aleurone cells and starchy endosperm cells in the cereal grain.

Richard D. Thompson is at the Max-Planck-Institut für Züchtungsforschung, Carl-von-Linne Weg 10, D-50829 Köln, Germany.
e-mail: thompson@mpiz-koeln.mpg.de

1. Becraft, P. W. & Asuncion-Crabb, Y. *Development* 127, 4039–4048 (2000).
2. Olsen, O.-A., Lemmon, B. & Brown, R. *Trends Plant Sci.* 3, 168–169 (1998).
3. Hueros, G., Varotto, S., Salamini, F. & Thompson, R. D. *Plant Cell* 7, 747–757 (1995).
4. Opsahl-Ferstad, H.-G., Le Deunff, E., Dumas, C. & Rogowsky, P. M. *Plant J.* 12, 235–246 (1997).



100 YEARS AGO

You occasionally do us, who are humble teachers of Elementary Science in schools, the very great kindness of giving us, through your columns, the chance of reaching the ears of those eminent men who are your frequent contributors, and who examine our pupils... I have spent a very large amount of time, money and labour in introducing the teaching of practical physics into my school, and trying to see that it shall be of the best kind possible, and I am prepared to do more. But really there must be some agreement between us and the said eminent men as to what practical science is when the examination paper is composed... The Cambridge Local Syndicate have introduced Elementary Experimental Science, three papers, into their junior syllabus. The other day I set two of these three papers for 1899 to a number of boys who had had a most careful experimental training in the matter of the syllabus... On marking the papers I found that the best boys... barely reached 40 per cent. of the marks. The same papers were set to a sharp boy of the same age who had done no experiments, but had been through the same subjects, mechanics, hydrostatics, and heat, in the old way, viz., text-book and problems. He scored nearly full marks on all the physics questions... Now these are exactly the old Cambridge — "Describe the common pump, &c., questions?" and the way to answer them is to waste no time on experiments, but read your text-book, get up your formulae and work examples.

From *Nature* 1 November 1900.

50 YEARS AGO

The aim of the Gordon Research Conferences is not so much to produce complete scientific papers, but rather to bridge the gap between advanced stages of the work and its appearance in print. How necessary this is in a rapidly developing field could be seen from the fact that, in the conference on ion exchange... certain fundamental investigations on ion-exchangers were reported by no less than five different research groups. These dealt with the swelling and exchange properties of resins of different degree of chain cross-linking; they showed uniformly that the uptake of swelling water and the speed of exchange decreases with increased cross-linking, while at the same time the specificity of the resin for different ions increases.

From *Nature* 4 November 1950.

Space physics

An auroral signature decoded

Patrick T. Newell

It has been hard to relate the activity of the northern lights to specific events in space. But the latest data show that 'auroral streamers' can be matched to bursts of ionized particles from Earth's magnetotail.

Aurorae — the northern and southern lights — occur high in the sky in polar regions. They result from the impact of energized electrons in near-Earth space with atoms and molecules in the upper atmosphere. It has been frustratingly difficult to ascribe particular auroral behaviour to events in near-Earth space. But four papers^{1–4}, published last month in *Geophysical Research Letters*, do just that.

Study of the near-Earth space environment originated with investigations of the aurora, which has remained central to space-physics research. This is in part because, in principle, the dynamics of ionized particles (plasma) in space are reflected in auroral behaviour. Just as a television screen glows when beams of electrons, accelerated by electric fields in near-vacuum conditions, strike its surface, so the upper atmosphere glows with shifting aurorae in response to accelerated beams of electrons from space.

This analogy is far from easy to apply in practice. But using coordinated data sets from spacecraft and ground observations made by the International Solar–Terrestrial Program (ISTP), the four papers^{1–4} identify a link between aurorae and events in the Earth's magnetotail — the long tail of plasma that extends far downstream from the Earth, on the night side facing away from the Sun (Fig. 1). They show that 'auroral streamers', which are fingers of light moving from the poleward edge of the aurora towards lower latitudes, exhibit a one-to-one correspondence with bulk flows of plasma that burst out of the magnetotail and are convected back towards the Earth. Debate is now underway as to whether these 'bursty bulk flows' are a key part of large auroral and magnetic disturbances. Such disturbances create the spectacular multicoloured curtains and rays of light associated with strong geomagnetic perturbations ('auroral substorms'), which, in extreme circumstances, can disrupt high-latitude communications, knock out power stations, and even harm satellites.

These developments in understanding aurorae owe much to the use of coordinated data sets. Key elements of the ISTP programme include the NASA Polar satellite, which can produce global images of the aurora in ultraviolet and X-ray wavelengths, and the Japanese Geotail satellite, which explores the magnetotail. Figure 2 shows one

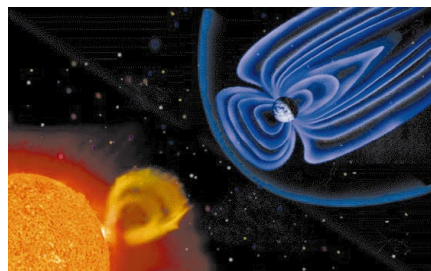


Figure 1 Sun, Earth and the magnetotail. An artist's view, not to scale, of the Sun and the magnetic bubble around Earth, the magnetosphere, which the Earth's magnetic field creates in the solar wind. On the side facing away from the Sun the magnetic field lines are stretched into a long magnetotail. (Image from ref. 8, courtesy of David Stern.)

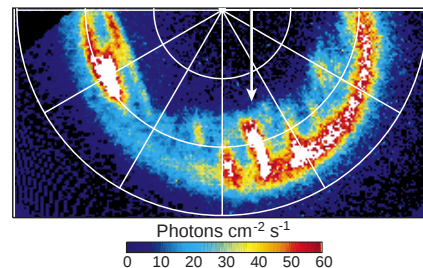


Figure 2 An auroral streamer, orientated north-south, indicated by the arrow. This ultraviolet image was taken by the Polar satellite. New work^{1–4} shows that auroral streamers stem from bursts of plasma that swirl back to the near-Earth space environment from the magnetotail.

auroral streamer as imaged by the Polar satellite. The Geotail satellite is the primary source of identifying bursty bulk flows, and there has been increasing evidence that such flows constitute most of the plasma and energy convection to the near-Earth region from the magnetotail.

Auroral streamers are a relatively little studied phenomenon. They are unusual in that, unlike most auroral activity, they extend north-south rather than east-west. They propagate southwards from the poleward (northern) edge of the oval-shaped torus of auroral activity (the 'auroral oval'). Because magnetic field lines closer to the pole extend out further into space, this direction of motion corresponds to plasma motion in the magnetotail towards Earth.

In 1998, Henderson *et al.*⁵ suggested

that bursty bulk flows of plasma from the magnetotail might correspond to auroral streamers. That work included no direct measurement of whether bursty bulk flows were actually occurring at the same time as auroral streamers. An association between the two came in work^{6,7} published last year, which combined global observations of the aurora from Polar with magnetotail flow data from Geotail. These results intrigued the space-physics community, although doubts remained — some even wondered whether a clear connection between auroral and magnetotail dynamics was theoretically possible.

In one of the new papers, Zesta *et al.*¹ use high-resolution ground-based measurements of the aurora to establish a detailed one-to-one correspondence between auroral streamers and the bursty bulk flows observed by Geotail. These results show that it is possible, while watching the aurora from the ground, to observe a specific behaviour (auroral streamers moving north-south) and associate it with a specific magnetotail phenomenon.

Also using ISTP data (from Geotail and a Canadian chain of ground-based observations called CANOPUS), Lui² shows that each new auroral streamer creates an auroral electrojet in the ionosphere. Auroral electrojets are electric currents driven longitudinally across the ionosphere. They are associated with auroral activity, and large ones may give rise to substorm magnetic activity lasting a few hours. Lui suggests that there is a continuum of behaviour, with full substorms at one end and isolated auroral streamers at the other.

Lyons and colleagues³ use solar-wind data from WIND (a NASA satellite, also part of ISTP), Geotail and CANOPUS to argue that, although auroral streamers are indeed caused by bursty bulk flows, the two are directly driven by variations in the solar wind. This idea runs counter to the more popular belief that such flows are part of internal magnetospheric dynamics. Finally, Anderson *et al.*⁴ use X-ray images of the aurora from Polar to conclude that changes in auroral luminosity are correlated with flows of plasma towards the Earth, but not particularly with the onset of substorms.

Auroral streamers are the first auroral signature to be clearly associated with a specific plasma process in the Earth's magnetotail. The bursty bulk flows that cause them may or may not prove to be an essential part of understanding intense auroral and magnetic disturbances. If they are, the long-debated trigger of auroral substorms may finally be understood. However that debate turns out, physicists can now decode at least one aspect of what aurorae are telling us. ■

Patrick T. Newell is at The Johns Hopkins University Applied Physics Laboratory, Johns Hopkins Road,

Laurel, Maryland 20723, USA.

e-mail: patrick.newell@jhuapl.edu

1. Zesta, E., Lyons, L. R. & Donovan, E. *Geophys. Res. Lett.* **27**, 3241–3244 (2000).
2. Lui, A. T. Y. *Geophys. Res. Lett.* **27**, 3245–3248 (2000).
3. Lyons, L. R., Zesta, E., Samson, J. C. & Reeves, G. D. *Geophys. Res. Lett.* **27**, 3237–3240 (2000).

4. Anderson, P. C., McKenzie, D. L., Lyons, L. R. & Hairston, M. *Geophys. Res. Lett.* **27**, 3233–3236 (2000).
5. Henderson, M. G., Reeves, G. D. & Murphree, J. S. *Geophys. Res. Lett.* **25**, 3737–3740 (1998).
6. Sergeev, V. A. *et al.* *Geophys. Res. Lett.* **26**, 417–420 (1999).
7. Fairfield, D. H. *et al.* *J. Geophys. Res.* **104**, 355–370 (1999).
8. <http://www-spf.gsfc.nasa.gov/Education/Intro.html>

Developmental biology

One cell, two fates

Peter Carmeliet

The two types of cell that make up blood vessels can develop from the same precursor. This discovery might improve our understanding of the role of blood vessels in disorders ranging from cancer to heart disease.

It seems odd, considering how important blood vessels are, how little we know about their development. We do know that they are generally composed of two cell types: endothelial cells line the inside, forming channels that conduct blood, while smooth muscle cells cover the outside, pro-

tecting the fragile channels from rupture and controlling blood flow. Conventional wisdom holds that endothelial and smooth muscle cells arise from separate 'precursor' cells through rounds of cell division and specialization.

But conventional wisdom is now being

overturned. On page 92 of this issue¹, Yamashita and colleagues describe a type of blood-vessel precursor from which both endothelial and smooth muscle cells develop (Fig. 1), in the culture dish and in mice. This precursor might contribute to the formation of both 'naked' endothelial capillaries and smooth-muscle-coated vessels — processes that are essential to development and reproduction and which, when derailed, contribute to cancer and some types of heart disease².

It was already known that, in the developing embryo, endothelial cells arise either from precursors — better known as 'progenitors' or 'stem cells' — that can produce only endothelial cells (angioblasts), or from progenitors that give rise to both endothelial and blood cells³ (haemangioblasts; Fig. 1). Endothelial precursors have also been found in adults^{4,5}. Most of these endothelial precursors divide and differentiate in response to a protein called vascular endothelial growth factor (VEGF), which is produced close to forming vessels^{4,5}.

The cells that surround the endothelial channels vary throughout the blood-vessel system⁶. For instance, several layers of smooth muscle cells surround the large vessels close to the heart, whereas single cells (not joined into layers) called pericytes cover smaller, more distant vessels.

These cells also differ in their origins (Fig. 1). The first smooth muscle cells in the embryo⁶ and the smooth-muscle-like cells in the prospective cardiac valves⁷ originate from the endothelium in a process that — at least in the heart — requires 'transforming growth factor' (TGF) proteins, such as TGF- β 3 (ref. 7). Another such protein (TGF- β 1) may be involved in the differentiation of the loose connective tissue (mesenchyme) around endothelial channels to form smooth muscle progenitors^{8,9}. Then, when these channels branch out, they produce a 'recruitment factor' — platelet-derived growth factor-BB (PDGF-BB). This instructs smooth muscle progenitors to migrate alongside the new branches^{8,9}. Endothelium-surrounding cells in the vessels that nourish the heart are derived from a putative progenitor in the external layer of the heart. By contrast, smooth muscle cells in the large vessels close to the heart and pericytes in the forebrain are derived from neural cells⁶. In adults, smooth muscle cells can form from many types of precursor¹⁰.

Yamashita *et al.*¹ further complicate this picture with their discovery that one type of embryonic precursor in the mouse can give rise to either endothelial cells or smooth muscle cells, depending on the growth factors to which it is exposed. In response to PDGF-BB, this precursor cell differentiates to form smooth muscle cells. These cells express a set of 'marker' molecules that distinguish smooth muscle cells from other

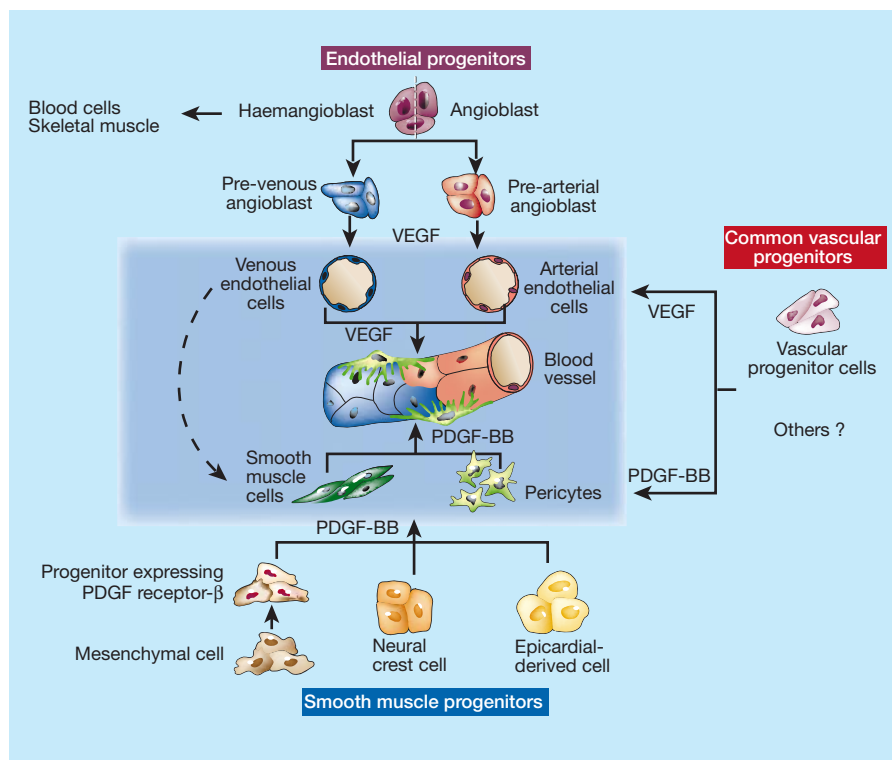


Figure 1 A common origin for the two types of blood-vessel cell. Centre, endothelial cells are those that make up the interior walls of a blood vessel. Smooth muscle cells or pericytes encase the endothelial channels. These basic cell types have been thought to arise from separate types of precursor. Endothelial cells arise from precursors called angioblasts or haemangioblasts in the embryo (top), or from circulating endothelial progenitors in the adult (not shown). Smooth muscle cells and pericytes, by contrast, can form from a variety of progenitors (bottom). These include mesenchymal cells (those that make up connective tissue) in embryo and adult, embryonic neural crest cells, and progenitors in the embryonic epicardium (the outer layer of the heart). Smooth muscle cells can also form from endothelial cells. But Yamashita *et al.*¹ have discovered a new precursor cell (right) that gives rise to both types of blood-vessel cell. Vascular endothelial growth factor (VEGF) promotes the development of endothelial cells from this precursor. Platelet-derived growth factor-BB (PDGF-BB) stimulates their development into smooth muscle cells and pericytes.

cells³, and surround endothelial channels in culture and *in vivo*.

VEGF, by contrast, sends the precursor along the developmental pathway to becoming an endothelial cell. But it seems that VEGF may only favour (and not fix) cell fate — which is not too surprising, given that endothelial cells still form in embryos lacking VEGF¹¹. Further molecules may fix endothelial fate. These molecules might resemble VEGF in binding one particular cellular receptor — VEGF receptor-2 — as the differentiation of endothelial cells is completely blocked when this receptor is lost¹². It will be important to identify these VEGF-like molecules, and to find out whether they affect Yamashita *et al.*'s versatile precursor cell. Another question is whether this vascular progenitor also specifies whether the endothelial cells it forms contribute to both arteries and veins (Fig. 1).

What about PDGF-BB — does it fix, or only favour, smooth-muscle-cell fate? Studies of mice engineered to lack PDGF-BB indicated that the mesenchymal cells mentioned above still develop into smooth muscle progenitors in the absence of PDGF-BB^{8,13}, as (to some extent) do Yamashita *et al.*'s common progenitor cells¹. But, in PDGF-BB-deficient mice, new branches from endothelial channels lack smooth muscle cells^{8,13}. It will be interesting to see whether this defect boils down to a problem with the growth and migration of the common vascular progenitors.

It is rather surprising that a common progenitor for endothelial and smooth muscle cells has only now been discovered. Precursors that give rise to several distinct types of neural cell, or to the different types of blood cell, have already been reported. And endothelial cells that line the inner surface of the heart share a common origin with their surrounding cardiac muscle fibres¹⁴. Moreover, endothelial cells can form skeletal muscle. Might there be other progenitor cells that have diverse fates in the blood-vessel system? And do haemangioblasts and Yamashita *et al.*'s vascular progenitors — which both express VEGF receptor-2 (ref. 3) — themselves arise from a single precursor?

Other unanswered questions relate to the possible medical implications of these results. Do the newly discovered progenitor cells, for instance, contribute to the growth of smooth muscle cells that occurs in atherosclerosis (the accumulation of lipids in, and thickening of, arterial walls)? Are they involved in the formation of new blood vessels that accompanies tumour development? Blindness can result when pericytes are lost from blood vessels in the retina — might it become possible to selectively shift these progenitors to become pericytes?

These vascular progenitors might also be useful to treat heart disease that is characterized by a lack of oxygen supply, as this requires

growth of both endothelial and smooth muscle cells³. In fact, signals from cancer cells, oxygen-depleted heart muscle and healing wounds in adults do recruit other endothelial progenitors^{4,5}. The next step is to learn more about how these versatile vascular precursors are induced to grow and differentiate. ■

Peter Carmeliet is at the Center for Transgene Technology and Gene Therapy, Katholieke Universiteit Leuven, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.
e-mail: peter.carmeliet@med.kuleuven.ac.be

1. Yamashita, J. *et al.* *Nature* **408**, 92–96 (2000).
2. Carmeliet, P. & Jain, R. K. *Nature* **407**, 249–257 (2000).

Nanotechnology

Flipping a molecular switch

Dan Feldheim

Switches lie at the heart of electronics and their design puts a limit on the size of integrated circuits. By harnessing chemistry, researchers have reduced this problem to a molecular level.

In the drive towards ever smaller electronic devices, many nanotechnologists are trying to exert chemical control over their components. Integrating molecular functions into this technology could lead to new types of electronic device. On page 67 of this issue Schiffrin and co-workers¹ take a first step towards synthesizing reversible molecular switches, and integrating these switches into a simple electronic circuit.

Take apart any digital microelectronic device and you will find basic electronic switches, known as transistors and diodes. More complex circuit elements, such as flip-flops, inverters or Schmitt triggers, are created from various permutations of these building blocks. For example, when two transistors are combined to form a Schmitt trigger, they create a circuit that changes from ON to OFF only if a voltage input exceeds a specific value. In this way, a Schmitt trigger defines a single bit of binary information (0 or 1). Schmitt triggers and flip-flop circuits are in turn combined to form more complex integrated circuits.

In their system, Schiffrin and co-workers switch their molecule's conductivity on and off by changing its oxidation state. The number of electrons associated with an atom determines its oxidation state. Some organic molecules contain a 'redox' centre where reduction (electrons added) or oxidation (electrons removed) occurs both readily and reversibly. If they are sandwiched as molecular layers between electrical contacts, such organic molecules can support relatively large currents². They do this through 'resonant tunnelling', a phenomenon that occurs when the electron energy levels of the molecule overlap with those of the metal it is in

3. Carmeliet, P. *Nature Med.* **6**, 389–395 (2000).
4. Rafii, S. *J. Clin. Invest.* **105**, 17–19 (2000).
5. Isner, J. M. & Asahara, T. *J. Clin. Invest.* **103**, 1231–1236 (1999).
6. Gittenberger-de Groot, A. C., DeRuiter, M. C., Bergwerff, M. & Poelmann, R. E. *Arterioscler. Thromb. Vasc. Biol.* **19**, 1589–1594 (1999).
7. Nakajima, Y., Mironov, V., Yamagishi, T., Nakamura, H. & Markwald, R. R. *Dev. Dyn.* **209**, 296–309 (1997).
8. Hellstrom, M. *et al.* *Development* **126**, 3047–3055 (1999).
9. Hirschi, K. K., Rohovsky, S. A. & D'Amore, P. A. *J. Cell Biol.* **141**, 805–814 (1998).
10. Campbell, J. H., Efendy, J. L., Han, C. L., Girjes, A. A. & Campbell, G. R. *J. Vasc. Res.* **37**, 364–371 (2000).
11. Carmeliet, P. *et al.* *Nature* **380**, 435–439 (1996).
12. Shalaby, F. *et al.* *Cell* **89**, 981–990 (1997).
13. Lindahl, P., Hellstrom, M., Kalen, M. & Betsholtz, C. *Curr. Opin. Lipidol.* **9**, 407–411 (1998).
14. Eisenberg, L. M. & Markwald, R. R. *Circ. Res.* **77**, 1–6 (1995).

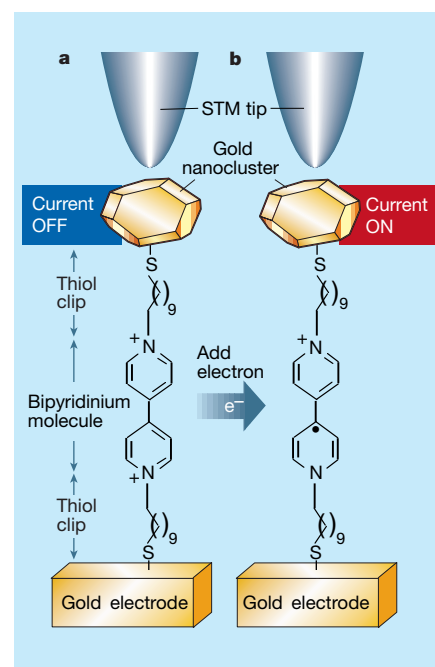


Figure 1 An electronic switch that works by changing the chemical state of a molecule. In the experiment by Schiffrin and co-workers¹, a bipyrindinium group (bipy) is attached to a gold electrode and a gold nanocluster by a chain of thiol 'clips'. The tip of a scanning tunnelling microscope (STM) is used to record the current flowing through the molecular switch. In a, the bipy molecule is in the oxidized bipy²⁺ state and no current flows. Electrons can be injected into the bipy molecule by applying a suitable voltage through the STM tip. b, When an electron is added the bipy molecule is reduced to bipy*⁺, and a large current flows. The size of this molecular switch is less than 10 nanometres.

contact with. This overlap is more likely to occur in molecules that are easily oxidized or reduced.

Schiffrin and co-workers¹ now show that a similar process can occur in a much smaller system. The authors have shown previously³ that a layer of an organic molecule containing the bipyridinium group (bipy) sandwiched between a gold electrode and gold nanoparticles can easily gain an electron and be reduced from the cation bipy^{2+} to the radical ion $\text{bipy}^{+\bullet}$. In their latest experiment, they attached an alkanethiol 'clip' to each end of a bipy molecule. One thiol was connected to the gold electrode and the other made contact with a gold nanocluster 6 nanometres in diameter (Fig. 1). The authors used the tip of a scanning tunnelling microscope to record the electrical properties of the nanoparticles individually. Their results show clearly how changing the redox state of the bipy molecules can control electron transport between the gold contacts. In other words, the molecules behave much like a molecular switch. The authors estimate that their 'molecular device' corresponds to no more than 60 organic molecules and needs fewer than 30 electrons to operate.

The way in which this new molecular device works is similar to the basic function of a Schmitt trigger. When the molecule is in the reduced $\text{bipy}^{+\bullet}$ state, relatively large currents flow through the nanocluster–molecule–electrode set-up characteristic of resonant tunnelling (Fig. 1). But when a certain threshold voltage is applied to the gold electrode, the tunnelling current decreases markedly. The threshold voltage corresponds to the oxidation of $\text{bipy}^{+\bullet}$ to bipy^{2+} .

Schiffrin and co-workers have created an electrochemical switch whose state is determined by the potential needed to reduce the molecule by one electron. Importantly, the threshold potential for ON/OFF switching should be tunable over a wide range simply by choosing molecules with easily reversible yet varying redox states. But will redox-switchable integrated nanoelectronic circuits ever be made? The answer is almost certainly no. A 'molecular' Schmitt trigger would operate too slowly and have insufficient gain (signal amplification) to be useful.

Nonetheless, there are possible applications for redox-switchable nanoelectronic devices where gain is not important^{2,4}. For example, chemical sensors based on the type of ligand-modified gold nanocluster described by Schiffrin and co-workers can, in theory, detect single molecules or single chemical reactions⁴. In addition, new forms of computing logic and memory are also being designed around redox-switchable molecular architectures^{4,5}.

The system reported by Schiffrin and colleagues adds to a rapidly expanding list of chemically or electrochemically switchable nanoelectronic devices. Earlier this year,

Heath and co-workers⁵ showed that the resistance of a monolayer of catenanes can be reversibly altered by several orders of magnitude using an electrochemically induced switching mechanism. Collins and co-workers⁶ showed that the electrical properties of carbon nanotubes can be switched chemically by dousing them with certain gases. Before these studies, work by my group had shown that single-electron tunnelling energies in individual gold nanoclusters can be controlled by adding protons to ligands bound to the cluster surface⁷.

Perhaps the most significant outcome of these studies will be a deeper fundamental understanding of basic relationships between structure and electronic function in nanoscale

structures. Future electronic devices may not contain flip-flops, or any other circuits we know now. By exploring the electrical properties of integrated molecular nanoelectronics, new electronic properties and applications are certain to emerge in the coming years. ■

Dan Feldheim is in the Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA.

e-mail: feldheim@chemdept.chem.ncsu.edu

1. Gittins, D. I., Bethell, D., Schiffrin, D. J. & Nichols, R. J. *Nature* **408**, 67–69 (2000).
2. Tao, N. J. *Phys. Rev. Lett.* **76**, 4066–4069 (1996).
3. Gittins, D. I. et al. *Adv. Mater.* **11**, 737–740 (1999).
4. Brousseau, L. C. III, Zhao, Q., Shultz, D. A. & Feldheim, D. L. *J. Am. Chem. Soc.* **120**, 7645–7646 (1998).
5. Collier, C. P. et al. *Science* **289**, 1172–1175 (2000).
6. Collins, P. G. et al. *Science* **287**, 1801–1804 (2000).
7. Feldheim, D. L. & Keating, C. D. *Chem. Soc. Rev.* **27**, 1–12 (1998).

Transcription

Regulation of the regulators

Scott Stewart and Gerald R. Crabtree

The list of proteins whose activity is controlled by transport into different subcellular compartments keeps growing. The latest additions are two regulators of gene expression.

Histone proteins serve double duty in the nucleus. First, they compress DNA and DNA-bound proteins into a compact structure called chromatin. Second, they have a say in which genes are expressed. Specifically, it is the way in which histones are labelled with different chemical groups that is important in controlling gene transcription. Acetyl (COCH_3) groups are one type of label, and are added and removed by enzymes called histone acetylases and deacetylases^{1,2}. These regulatory enzymes must themselves be kept under control —

but how? By looking at how muscle cells differentiate to form muscle fibres in cell culture, McKinsey and colleagues (page 106 of this issue³) provide some answers. Unusually, they find that two types of histone deacetylase are moved out of the nucleus during muscle differentiation, in a process that ultimately is controlled by calcium ions — ubiquitous chemical messengers.

The process of converting an unspecialized muscle cell to a skeletal muscle fibre (myogenesis) is a complicated business. The key point is that an array of muscle-specific

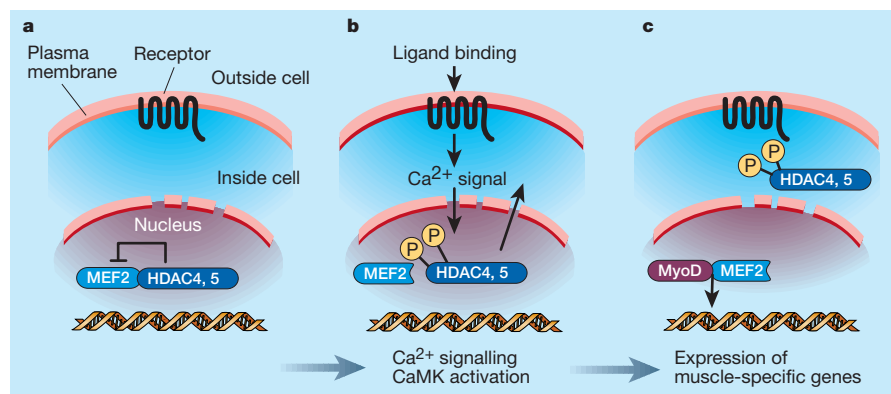


Figure 1 One way of controlling the regulators of gene transcription. Histone deacetylase (HDAC) enzymes regulate transcription by removing acetyl groups from histone proteins bound to DNA (not shown). a, In cultured early muscle cells, HDAC4 and HDAC5 bind to a muscle-specific transcription factor, MEF2, and inhibit the transcription of muscle-specific genes. b, McKinsey *et al.*³ find that removal of serum from the culture medium initiates a calcium signal (possibly when a ligand such as insulin-like growth factor-1 binds to its receptor on the cell surface). Calcium activates calcium/calmodulin-dependent protein kinases (CaMKs) I and IV (not shown). These enzymes add phosphate (P) groups to HDAC4 and HDAC5, which then dissociate from MEF2 and are ejected from the nucleus. c, MEF2 is then free to bind to transcription factors such as MyoD and activate the expression of muscle-specific genes.

genes (such as that encoding the myosin protein, a molecular motor) must be turned on. This requires the cooperation of several proteins⁴ including the transcription factor MyoD and myocyte enhancer factor-2 (MEF2).

Clearly, the activity of these factors must be controlled so they do not wantonly activate muscle-specific genes. This is where the histone deacetylases come in. Several groups have reported that MEF2 and histone deacetylases interact functionally and physically^{5–9}, resulting *in vitro* in the repression of MEF2. In both T cells and early muscle cells in culture, MEF2 responds to Ca^{2+} signals by dissociating from histone deacetylases^{7,10}. Moreover, histone deacetylase-4 can shuttle into and out of the nucleus⁵, and histone deacetylase-4 and -5 can be kept outside the nucleus by proteins of the 14-3-3 family¹¹. McKinsey *et al.*'s triumph is to link these seemingly disparate observations, and to show their relevance to muscle differentiation and gene expression³. Their results also provide the first example of the 'spatial' regulation (by export from the nucleus) of a chromatin regulator.

The authors report that when early muscle cells are encouraged to differentiate by the removal of serum from their culture medium, histone deacetylase-5 exits the nucleus. MEF2 is left behind, where it promotes the transcription of genes essential for muscle differentiation. This response is thought to be mediated largely by Ca^{2+} signals generated at the cell surface. In light of this fact and earlier results⁷, McKinsey *et al.* set out to find a Ca^{2+} -regulated enzyme that might induce the removal of histone deacetylase-5 from the nucleus. The answer turned out to be two Ca^{2+} /calmodulin-dependent protein kinases (enzymes that add phosphate groups to other proteins). When added to muscle cells in culture, the constitutively active form of each of these enzymes (known colloquially as CaMKI and CaMKIV) induces the export of added histone deacetylase-5 (and -4) from the nucleus. The response is selective to histone deacetylase-4 and -5 — added CaMKI did not have the same effect on histone deacetylase-1 — and seems to require the phosphorylation of the histone deacetylases in question.

Previous work⁷ showed that only Ca^{2+} and a nuclear extract are required to induce the dissociation of histone deacetylase-4 from MEF2. So, putting all these results together, we can suggest how the pathway in early muscle cells might work (Fig. 1). Binding of an extracellular molecule to its receptor on the cell surface leads to an increase in cellular Ca^{2+} . This then activates Ca^{2+} /calmodulin-dependent kinases, which in turn phosphorylate histone deacetylase-4 and -5. These enzymes then detach themselves from MEF2, exposing their nuclear export signal (also discovered by McKinsey *et al.*). This signal consists of a sequence of

amino acids that acts as an address label, allowing the proteins to be sent out of the nucleus. Alone in the nucleus, MEF2 works together with muscle-specific transcription factors to enhance the expression of muscle-specific genes, allowing differentiation of the cell into a muscle fibre. This all seems physiologically plausible, especially as the repression of MEF2 by a constitutively nuclear histone deacetylase-5, with mutations in its sites of phosphorylation by CaMKI, cannot be relieved by CaMKI (or, apparently, by Ca^{2+} , although this was not shown directly).

This is all well and good, but there are still details of the signalling pathway to be resolved. The most unintuitive part of the pathway is that removal of serum induces a Ca^{2+} -dependent process. Serum contents such as platelet-derived growth factor normally stimulate cells, so one might think that removing serum would terminate stimulation. One possible explanation is that a molecule that is activated by serum withdrawal is important — perhaps insulin-like growth factor-1, which induces muscle differentiation¹², or phenylephrine, which invokes Ca^{2+} signalling and the dissociation of histone deacetylases from MEF2 (ref. 9). However, McKinsey *et al.*'s results³ do not establish that a Ca^{2+} stimulus alone is sufficient for histone deacetylase-4 and -5 to exit the nucleus. This question could be easily answered by using drugs, such as ionomycin, that cause selective Ca^{2+} influx at physiological levels and rates.

Another question is how relevant these mechanisms are to muscle differentiation *in vivo*. This may become clear only with the use

of mice in which mutant histone deacetylase genes, encoding mutations in the phosphorylation sites, are inserted into the mice's own histone deacetylase gene. McKinsey *et al.*'s results predict that the histone deacetylase encoded by such mutant genes would not be able to respond to Ca^{2+} signalling. It might therefore block the differentiation of skeletal muscle (and, likewise, the programmed cell death of T cells, during which MEF2 also dissociates from histone deacetylases). But whatever the outcome of these *in vivo* studies, the new work³ is important not least because it adds to the growing body of evidence¹³ for the regulation of chromatin regulators by signals that begin at the cell membrane.

Scott Stewart and Gerald R. Crabtree are in the Departments of Developmental Biology and Pathology, Howard Hughes Medical Institute, Stanford University Medical School, 279 Campus Drive, Stanford, California 94305, USA.
e-mail: crabtree@cmgm.stanford.edu

1. Brownell, J. E. *et al.* *Cell* **84**, 843–851 (1996).
2. Taunton, J., Hassig, C. A. & Schreiber, S. L. *Science* **272**, 408–411 (1996).
3. McKinsey, T. A., Zhang, C.-L., Lu, J. & Olson, E. N. *Nature* **408**, 106–111 (2000).
4. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51**, 987–1000 (1987).
5. Miska, E. A. *et al.* *EMBO J.* **18**, 5099–5107 (1999).
6. Wang, A. H. *et al.* *Mol. Cell. Biol.* **19**, 7816–7827 (1999).
7. Youn, H. D., Grozinger, C. M. & Liu, J. O. *J. Biol. Chem.* **275**, 22563–22567 (2000).
8. Lemerrier, C. *et al.* *J. Biol. Chem.* **275**, 15594–15599 (2000).
9. Lu, J. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 4070–4075 (2000).
10. Youn, H. D., Sun, L., Prywes, R. & Liu, J. O. *Science* **286**, 790–793 (1999).
11. Grozinger, C. M. & Schreiber, S. L. *Proc. Natl Acad. Sci. USA* **97**, 7835–7840 (2000).
12. Musarò, A. *et al.* *Nature* **400**, 581–585 (1999).
13. Zhao, K. *et al.* *Cell* **95**, 625–636 (1998).

Astronomy

Cosmic discord

Craig J. Hogan

New data on the cosmic background radiation are making cosmologists revise their view of the Universe. The biggest surprise is that the data favour a larger amount of 'ordinary matter' than was thought.

The early Universe is a simple system composed of nearly uniform ordinary matter, radiation and dark matter. Cosmologists say that an early period of rapidly accelerating expansion (inflation) created small-scale fluctuations that grew into the large and complicated structures we see today (galaxies, stars and planets). Larger-scale fluctuations are also imprinted on the cosmic microwave background, or CMB — the faint radiation left over from the Big Bang. Recent measurements of temperature variation in the CMB, especially from the Boomerang and Maxima experiments^{1,2}, reveal distinctive patterns in these fluctuations, which depend on the details of the

shape and composition of the Universe.

In several papers appearing on the web^{3–7} and now in print⁸, cosmologists are starting to interpret these patterns through detailed statistical studies. They are finding amazing consistency with the expectations of inflation — a nearly 'flat' Universe, which can be described as a small piece of an enormous hypersphere — and with independent estimates of quantities such as the densities of dark matter and 'dark energy'. But there are also a few unexpected and possibly important discrepancies.

The sharpest and most interesting discrepancy is the estimate of the density of ordinary (baryonic) matter in the Universe: the new data say that the mean number of

neutrons and protons per unit volume is greater than was thought. The baryon density connects with several wildly different observations of the Universe. The mean density of baryons when the Universe was 300,000 years old affects the properties of primordial acoustic waves, which directly influence observable fluctuations of the CMB measured by Boomerang and Maxima. At much earlier times, at a cosmic age of about one second to a few minutes, the baryon density is the one parameter that determines the production of light nuclei — especially the fractions of deuterium (D), helium (He) and lithium (Li), whose primordial abundances are estimated from observations of stars and gas more than a billion years later^{9,10}. Finally, we can estimate the mean density of baryons simply by looking around today and counting them — most of them are found in hot, ionized plasma on the outskirts of galaxies¹¹.

Recent improvements in these three estimates have produced remarkable concordance. But close examination of the best current data now shows a statistically significant discrepancy: the CMB data imply a higher baryon density than the abundance of light nuclei does, and a reasonable amount of waffling in other cosmological parameters seems unable to eliminate the problem. It seems that we are bound to learn something new — perhaps about the interpretation of one of the data sets or one of the other cosmological parameters, or perhaps about the standard cosmological model, which is used to weave together the different observations (Box 1).

The first map of the CMB, from 1992, revealed large-scale fluctuations (at a scale of 10° to 90° across the sky), but Boomerang and Maxima can measure much smaller-angle temperature fluctuations. They find the expected broad 'acoustic peak' in the power spectrum at an angular scale of around 1°, followed by smaller peaks at even smaller scales. The relative amplitudes and positions of these peaks are sensitive to the baryon density because the baryons contribute the main inertia in the primordial acoustic waves responsible for the pattern of

peaks. The current data conflict with the predictions of the standard cosmological model because the smaller peaks, especially the second acoustic peak, are much weaker than expected. By varying the favoured values of the cosmological parameters within the constraints of the data to explore the range of viable universes, it seems that no simple statistical fix can resolve the discrepancy⁸.

Immediate suspicion falls on the baryon density. A higher baryon density increases the amplitude of the odd peaks (the first and third) relative to the even ones (the second and fourth). If the baryon density is made a factor of two higher than the value indicated by the abundance of light nuclei, the second acoustic peak would be lowered in line with the Boomerang and Maxima data. Such an adjustment seems reasonable, because considerable interpretation is required to derive the baryon density from abundance data^{9,10}.

Nonetheless, the baryon density preferred by the CMB data predicts nuclear abundances that are unacceptable to anyone. The best fit to the Boomerang data⁴ using the favoured Hubble constant of $H_0 = 71 \pm 8$ (which gives the expansion rate of the Universe in $\text{km s}^{-1} \text{Mpc}^{-1}$), estimates a baryon density, Ω_b , of $7.4 \pm 1\%$. This translates to a baryon/photon ratio (in dimensionless units of 10^{-9}), $\eta_{-9} = 1.0 \pm 0.15$. For $\eta_{-9} = 1$, standard models of nucleosynthesis during the Big Bang yield the following abundances for the light elements relative to hydrogen: $\text{D}/\text{H} = 1.12 \times 10^{-5}$, ${}^7\text{Li}/\text{H} = 1.0 \times 10^{-9}$ and a He abundance of 25.2% by mass.

These estimates do not agree with the available observations. The predicted D/H is lower than direct measurements of abundances in quasar absorbers, the interstellar medium, and the jovian atmosphere — a paradox because the Big Bang is thought to be the unique cosmic source for deuterium and subsequent evolution destroys it. The He abundance by mass is only about 1–2% above the observational estimates in metal-poor dwarf galaxies, but this is still thought to exceed the systematic measurement error. Current estimates of Li from metal-poor

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

stars show a 'plateau' at $\text{Li}/\text{H} \sim 10^{-10}$ (ref. 12), which may be close to the primordial abundance. Although it might be possible to uniformly destroy 90% of the ${}^7\text{Li}$ these stars started with, there are good arguments against this option — including the presence of ${}^6\text{Li}$, which is far more fragile.

There are other ways to remove the baryon-density discrepancy, but none of them is compelling. There may, for example, be a new source of reionization at very high redshift that creates a thicker electron-scattering photosphere through which we view the fluctuations. There might be a 'tilt' in the primordial spectrum that changes the initial angular dependence (although this would spoil the beautiful agreement of intermediate-scale fluctuations with both larger-scale fluctuations and galaxy clustering). Or other sources of dissipation could be introduced that would damp acoustic waves more quickly than standard theory predicts. Some of these possibilities can be ruled out by examining the smaller acoustic peaks. The tilt and damping options lead to all the smaller peaks being suppressed, whereas a higher baryon density predicts a return to a higher value for the third peak, as well as a shift in the location of the peaks owing to the difference in sound speed.

It is surprising that our various measures of baryon density agree so well. After all, we are comparing evidence from processes that operate at cosmic ages and scales differing by more than ten orders of magnitude, and find values for the baryon density that agree to within a factor of two. Nonetheless, exploring this discrepancy might lead to something really new — perhaps a simple reinterpretation of data on abundances, or perhaps a new ingredient not yet included in the standard cosmological model. ■

Craig J. Hogan is in the Department of Astronomy, University of Washington, Seattle, Washington 98195, USA.

e-mail: hogan@astro.washington.edu

Box 1 The concordance model

Only about seven or eight parameters are needed to specify a simple cosmological model. They are: the rate of expansion of the Universe today (the Hubble constant, H_0 in units of $\text{km s}^{-1} \text{Mpc}^{-1}$); the mean curvature of space, K ; the densities of several forms of matter (radiation, baryons, collisionless dark matter and a cosmological constant are popular choices, although only

the first two have actually been seen directly); and two perturbation parameters that determine amplitudes of radiation anisotropy and the current clustering of matter. These parameters are enough to specify the current global state of the Universe on large scales; with general relativity, they are enough to extrapolate its behaviour back to within microseconds of the Big Bang.

The favoured 'concordance' model before the latest CMB measurements had a Universe with 3–5% baryons, 30% dark matter, 65% vacuum density (cosmological constant) and $H_0 \sim 70$. The new Boomerang and Maxima results appear to imply a much higher baryon density, around 7%, which is inconsistent with the pattern of light-element abundances. **C.J.H.**

1. De Bernardis, P. *et al.* *Nature* **404**, 955–959 (2000).
2. Hanany, S. *et al.* <http://xxx.lanl.gov/abs/astro-ph/0005123>
3. Lange, A. E. *et al.* <http://xxx.lanl.gov/abs/astro-ph/0005004>
4. Balbi, A. *et al.* <http://xxx.lanl.gov/abs/astro-ph/0005124>
5. Hu, W., Fukugita, M., Zaldarriaga, M. & Tegmark, M. <http://xxx.lanl.gov/abs/astro-ph/0006436>
6. Lesgourgues, J. & Peloso, M. <http://xxx.lanl.gov/abs/astro-ph/0004412>
7. White, M., Scott, D. & Pierpaoli, E. <http://xxx.lanl.gov/abs/astro-ph/0004385>
8. Tegmark, M. & Zaldarriaga, M. *Phys. Rev. Lett.* **85**, 2240–2243 (2000).
9. Nollett, K. M. & Burles, S. *Phys. Rev. D* **61**, 123505 (2000).
10. Burles, S., Nollett, K. M. & Turner, M. S. <http://xxx.lanl.gov/abs/astro-ph/0008495>
11. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. *Astrophys. J.* **508**, 518–530 (1998).
12. Ryan, S. G. *et al.* *Astrophys. J.* **523**, 654–677 (1999).

Global spread of microorganisms by ships

Ballast water discharged from vessels harbours a cocktail of potential pathogens.

Commercial ships have spread many species around the world^{1–3}, but little is known of the extent and potential significance of ship-mediated transfer of microorganisms^{3,4}. Here we show that the global movement of ballast water by ships creates a long-distance dispersal mechanism for human pathogens and may be important in the worldwide distribution of microorganisms, as well as for the epidemiology of waterborne diseases affecting plants and animals.

Ships have used ballast water for stability since the nineteenth century, discharging water at ports of call and en route¹. Ports can receive relatively large volumes of ballast water — for example, the United States receives more than 79 million tonnes of ballast water from overseas each year⁵. Ballast tanks carry a diverse community of organisms, resulting in many biological invasions^{2,3}. Pathogens, including those affecting humans, are common in coastal waters⁶ and can also be transferred in ballast water^{7,8}.

We measured the concentrations of total bacteria, virus-like particles (VLPs) and the bacteria *Vibrio cholerae* O1 and O139, which cause human epidemic cholera, in the ballast water of vessels arriving to Chesapeake Bay (Fig. 1) from foreign ports. We collected water samples to estimate the abundance of total bacteria and VLPs, and also took both water and plankton samples to measure the concentration of *V. cholerae*, as the bacterium forms associations with plankton⁹.

Our samples contained an average of 8.3×10^8 bacteria (s.e., 1.7×10^8 ; $n=11$) and 7.4×10^9 VLPs (s.e., 2.3×10^9 ; $n=7$) per litre. Given that Chesapeake Bay received an estimated 1.2×10^{10} litres of foreign ballast water in 1991 alone⁵, our measures indicate that ballast water probably delivers large numbers of microbial species and potential pathogens to this estuary.

Vibrio cholerae was found in plankton samples from all ships, and both serotypes were detected in 93% of the ships (Fig. 2a). The concentration of *V. cholerae* O1 was significantly greater than that of O139 (Fig. 2a; Student's *t*-test, $t=2.296$; d.f., 27; $P<0.05$). Furthermore, there were 100 times more *V. cholerae* O1 and O139 in water samples than in plankton samples from the same ships (Fig. 2b; paired *t*-tests for serotype O1 and O139 respectively, $t=8.10$; d.f., 6; $P<0.001$ and $t=10.05$; d.f., 6; $P<0.001$), indicating that *V. cholerae* was not concentrated on zooplankton larger than 80 μm in diameter.

Vibrio cholerae is a useful model to examine the possible significance of ballast-



Figure 1 Chesapeake Bay, on the US East Coast, receives some ten billion litres of foreign ballast water each year. Each litre typically contains about a billion bacteria and seven billion virus-like particles.

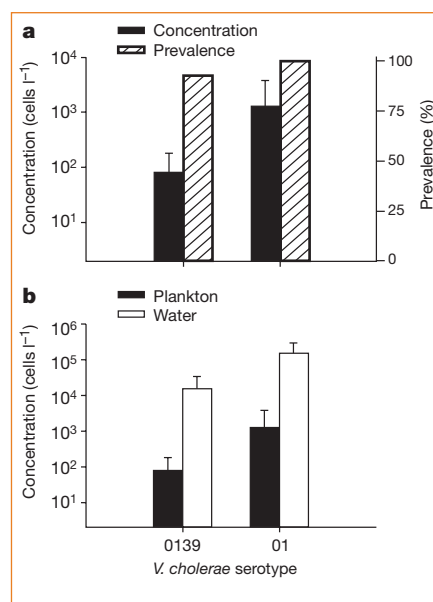


Figure 2 Prevalence and concentration of *Vibrio cholerae* serotype O1 and O139 in ships' ballast water. **a**, Prevalence and concentration associated with plankton samples. Prevalence shows the percentage of ships in which the respective serotypes were detected. Sample size differed between serotype O1 ($n=15$ ships) and O139 ($n=14$ ships). **b**, Comparison of concentration of each serotype between paired plankton and water samples from the same ballast tanks ($n=7$ ships). Concentrations are shown as mean and standard error for respective serotypes and sample types as detected by direct count methods using fluorescent antibodies; further details are available from the authors.

mediated dispersal in transmission of pathogens. Our data indicate that *V. cholerae* can be delivered frequently by ships to estuaries with commercial ports, and our observations of dividing cells revealed that some bacteria are viable upon arrival. Although it remains difficult to estimate the concentration of viable cells¹⁰, the transfer and release of *V. cholerae* by ships creates an opportunity for the colonization of coastal ecosystems. *V. cholerae* is a common component of freshwater and marine habitats,

including Chesapeake Bay, where it persists without human contact^{9,10}. Thus, should a novel genotype arrive in ballast water, local conditions may favour its establishment. The extent to which this has occurred, and the degree of geographic differences in the genetic structure of *V. cholerae* populations, is unknown.

We predict that coastal ecosystems are frequently invaded by microorganisms from ballast water. First, concentrations of bacteria and viruses exceed those reported for other taxonomic groups in ballast water by 6–8 orders of magnitude², and the probability of successful invasion should increase with inoculation concentration¹¹. Second, the biology of many microorganisms may facilitate invasion, combining a high capacity for increase, asexual reproduction, and the ability to form dormant resting stages^{3,12}. Such flexibility in life history can broaden the opportunity for successful colonization, allowing rapid population growth when suitable environmental conditions occur. Third, many microorganisms can tolerate a broad range of environmental conditions, such as in salinity or temperature, so many sites may be suitable for colonization^{6,12}. This suite of factors may yield a high rate of invasion for microorganisms compared to invertebrates, which are already known to invade coastal habitats from ballast water^{2,3}.

Despite growing concern about biological invasions^{7,11} and emergent diseases^{9,13,14}, the extent and effects of the transfer of microorganisms in ballast water are virtually unexplored³. We know of no published estimates of microbial genetic diversity in ballast water, and the fate of microorganisms discharged from ballast tanks remains unknown³. Given the magnitude of ongoing transfer and its potential consequences for ecological and disease processes, large-scale movement of microorganisms by ships merits attention from both invasion biologists and epidemiologists.

Gregory M. Ruiz*, **Tonya K. Rawlings*†**,
Fred C. Dobbs‡, **Lisa A. Drake‡**,
Timothy Mullady*, **Anwarul Huq†**,
Rita R. Colwell†

*Smithsonian Environmental Research Center,
PO Box 28, Edgewater, Maryland 21037, USA
e-mail: ruiz@serc.si.edu

†Center of Marine Biotechnology, University of
Maryland Biotechnology Institute, 701 East Pratt
Street, Baltimore, Maryland 21202, and Department
of Cell and Molecular Biology, University of
Maryland, College Park, Maryland 20742, USA

‡Department of Ocean, Earth and Atmospheric
Sciences, Old Dominion University, Norfolk,
Virginia 23529, USA

1. Carlton, J. T. *Oceanogr. Mar. Biol. Annu. Rev.* **23**, 313–374 (1985).
2. Carlton, J. T. & Geller, J. B. *Science* **261**, 78–82 (1993).
3. Ruiz, G. M., Fofonoff, P. W., Carlton, J. T., Wonham, M. J. & Hines, A. H. *Annu. Rev. Ecol. Syst.* (in the press).
4. Galil, B. S. & Hülsmann, N. *Eur. J. Protistol.* **33**, 244–253 (1997).
5. Carlton, J. T., Reid, D. & van Leeuwen, H. *The Role of Shipping in the Introduction of Nonindigenous Aquatic Organisms to the Coastal Waters of the United States (other than the Great Lakes) and an Analysis of Control Options* (Tech. Rep. No. CG-D11-95, US Coast Guard, Washington DC, 1995).
6. Grimes, D. J. *Estuaries* **14**, 345–360 (1991).
7. Hallegraeff, G. M. *Mar. Ecol. Prog. Ser.* **168**, 297–309 (1998).
8. McCarthy, S. A. & Khambaty, F. M. *Appl. Environ. Microbiol.* **60**, 2597–2601 (1994).
9. Colwell, R. R. *Science* **274**, 2025–2031 (1996).
10. Colwell, R. R. & Huq, A. in *Vibrio cholerae and Cholera: Molecular to Global Perspectives* (eds Wachsmuth, I., Blake, P. & Olsvik, Ø.) 117–134 (Am. Soc. Microbiology, Washington DC, 1994).
11. Williamson, M. *Biological Invasions* (Chapman & Hall, London, 1996).
12. Brock, T. D., Madigan, M. T., Martinko, J. M. & Parker, J. *Biology of Microorganisms* 7th edn (Prentice Hall, Englewood Cliffs, NJ, 1993).
13. Harvell, C. D. *et al. Science* **285**, 1505–1510 (1999).
14. Daszak, P., Cunningham, A. A. & Hyatt, A. D. *Science* **287**, 443–449 (2000).

Materials science

The smallest carbon nanotube

We report here the discovery of the smallest possible carbon nanotube. This has a diameter of 4 Å, which is the narrowest attainable that can still remain energetically stable, as predicted by theory. These nanotubes are confined inside multiwalled carbon nanotubes and their diameter corresponds to that of a C₂₀ dodecahedron with a single carbon atom at each of its twenty apices. Unlike larger carbon nanotubes, which, depending on their diameter and helicity, can be either metallic or semiconducting, these smallest nanotubes are always metallic.

Many of the extraordinary properties of carbon nanotubes¹ depend on their diameter. The smallest carbon nanotubes have been associated with the smallest fullerenes², with nanotube diameters of 7 or 5 Å corresponding to those of C₆₀ and C₃₆ structures, respectively. Carbon nanotubes with such small diameters have been observed experimentally^{3–5}.

The carbon nanotubes reported here were seen in cathodic deposits produced by arc-discharge of graphite rods in a hydrogen atmosphere without a metallic catalyst⁶. Under these conditions, more carbon nanotubes (all multiwalled) are kept open as hydrogen etches away the capping atoms, a unique feature that helps maintain a favourable environment for smaller carbon nanotubes to form inside already-grown multiwalled carbon nanotubes.

Figure 1 shows a high-resolution transmission electron micrograph of an 18-shell carbon nanotube produced in this way: each dark line represents the side wall of a cylindrical graphene shell in projection and the innermost shell has a diameter of 4 Å. The cylindrical structure of the nanotube is shown by the reduced contrast towards the centre of the nanotube, where there are fewer atoms in the smaller tubes. Multiwalled carbon nanotubes with innermost shells of diameter 5 or 7 Å were also detected in the same material.

Some of the smallest nanotubes seen were capped, and electron micrographs of the cap of a 4 Å nanotube indicate that the smallest nanotubes have an antichiral [3, 3] ‘armchair’ structure⁷ (Fig. 1, inset). We propose that these 4 Å nanotubes grow out of half of a C₂₀ dodecahedron, in which the C–C bonding angle is 108° — very close to the bonding angle (109.5°) in the sp³ configuration of diamond. Growth into a full nanotube is realized through a step-by-step mechanism on both the outer and inner surfaces⁸ (Fig. 1, inset). The hydrogen atmosphere facilitates the formation of the



Figure 1 High-resolution electron microscope image of a 4 Å tubule (side walls are marked by lines) confined inside an 18-shell carbon nanotube. Dark lines correspond to the side walls of the single-wall shells. The blurring towards the centre is due to the smaller number of atoms in the smaller tubules. Inset, model for the formation and growth of a [3,3] armchair tubule of 4 Å diameter. Half of the C₂₀ dodecahedron (yellow) serves as a nucleus and further growth into the tubule is realized by adding carbon atoms (blue).

halves of the C₂₀ dodecahedron, stabilizing the structure by terminating certain dangling bonds with hydrogen, and keeps the inside of the nanotubes open so that, after the confining outer shells have been formed, carbon species can enter the core to form the innermost shell.

Although these tiny 4 Å carbon nanotubes should be energetically stable⁹, the severe steric distortion resulting from the planar graphene changes the electronic structure significantly¹⁰. Electronic band-structure calculations indicate that all such tubules tend to be metallic, regardless of their helicity.

C₂₀ fullerenes have recently been made using a gas-phase reaction method¹¹. Our finding indicates that C₂₀ fullerenes could also be produced by the arc-discharge method. It remains a challenge to produce single-walled carbon nanotubes of 4 Å diameter experimentally.

Lu-Chang Qin*, Xinluo Zhao†,

Kaori Hirahara‡, Yoshiyuki Miyamoto‡, Yoshinori Ando§, Sumio Iijima*†‡§

*JST-ICORP Nanotubulite Project, †NEC Corporation, 34 Miyukigaoka, Tsukuba 305-8501, Japan

email: qin@frl.cl.nec.co.jp

‡JST-ICORP Nanotubulite Project, §Department of Materials Science and Engineering, Meijo University, Tempaku-ku, Nagoya 468-8502, Japan

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Heath, J. R. *Nature* **393**, 730–731 (1998).
3. Ajayan, P. M. & Iijima, S. *Nature* **358**, 23 (1992).
4. Kataura, H., Achiba, Y., Zhao, X. & Ando, Y. *Mat. Res. Soc. Symp. Proc.* **593**, 113–118 (2000).
5. Sun, L. F. *et al. Nature* **403**, 384 (2000).
6. Zhao, X. *et al. Carbon* **35**, 775–781 (1997).
7. Qin, L.-C. *et al. Chem. Phys. Lett.* **268**, 101–106 (1997).
8. Iijima, S., Ajayan, P. M. & Ichihashi, T. *Phys. Rev. Lett.* **69**, 3100–3103 (1992).
9. Sawada, S. & Hamada, N. *Solid State Commun.* **83**, 917–919 (1992).
10. Blase, X., Benedict, L. X., Shirley, E. L. & Louie, S. G. *Phys. Rev. Lett.* **72**, 1878–1881 (1994).
11. Prinzbach, H. *et al. Nature* **407**, 60–63 (2000).

Materials science

Single-walled 4 Å carbon nanotube arrays

Here we describe the smallest carbon nanotubes possible¹, prepared by the pyrolysis of tripropylamine molecules in the channels of porous zeolite AlPO₄-5 (AFI) single crystals². These uniformly sized carbon nanotubes have a diameter of 0.4 nm and are the best example of one-dimensional quantum wires.

AFI is a type of transparent microporous crystal (Fig. 1) containing one-dimensional channels packed in hexagonal arrays, with an inner diameter of 0.73 ± 0.01 nm (ref. 3). The starting material we used for synthesizing single-walled carbon nanotubes (SWNTs) was tripropyl-

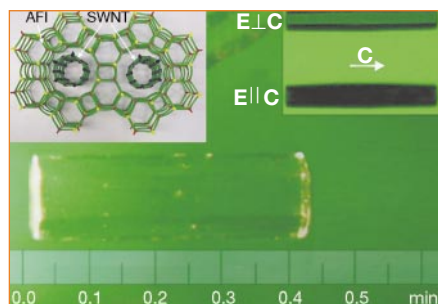


Figure 1 An as-grown zeolite $\text{AlPO}_4\text{-5}$ (AFI) single crystal. Left inset, model of the AFI crystal structure showing single-walled nanotubes (SWNTs) inside the channels. Right inset, polarization characteristics of the SWNT-AFI sample: no absorption (top) and high absorption (bottom) for light polarized perpendicular ($E_{\perp}$) and parallel ($E_{\parallel}$) to the tube direction, respectively. E , electric field of polarized light; C , zeolite channel direction.

amine, introduced into the channels during growth of AFI. SWNTs form inside the AFI channels when the pyrolysed carbon is thermally treated² (Fig. 1). The resulting brown AFI crystallites have good polarization characteristics (Fig. 1, inset). Polarized Raman spectra of these AFI crystals match well with theoretically predicted Raman-active modes, including the breathing modes^{4,5}.

To view the SWNTs in a transmission electron microscope (TEM), we first dissolved the AFI framework in 30% hydrochloric acid; the SWNT-containing solution was then enriched and dispersed on a lacey carbon film for high-resolution TEM (JEOL2010 microscope operating at 200 kV). Figure 2 shows a typical image in which the specimen is represented mainly by two types of morphology: small pieces of highly curved, raft-like graphite stripes ('G' in Fig. 2) and ultrathin SWNTs ('T' and large arrowheads in Fig. 2). In the 'G' areas, the spacing between neighbouring parallel

fringes is measured to be about 0.34 nm, which matches well with the {002} lattice plane spacing of graphite. The SWNTs show a typical contrast effect, with the tube walls appearing as paired dark fringes; this contrast effect is very weak because of the small dimension of the tubes.

By using the {002} spacing of graphite as an internal reference and measuring the separation between the paired dark fringes of the SWNT images, we determined the diameters of the SWNTs to be 0.42 ± 0.02 nm. In the image shown in Fig. 2, there are ten or more SWNTs visible, all having the same morphology. The SWNTs in our samples were unstable under electron beam radiation, fading during observation in 10–15 seconds, whereas the raft-like layers of graphite persisted. We believe that the large number of small graphite pieces, which generally contain several layers, are transformed from SWNTs during extraction from the AFI crystal.

The narrowest diameter reported in the centre of multiwalled nanotubes is 0.5 nm (ref. 6), the same as that of a C_{36} molecule⁷. The smallest free-standing SWNT, of diameter 0.7 nm (the same as that of C_{60}), is stable⁸. The SWNTs observed here, which are consistent with a diameter of about 0.4 nm, are to our knowledge the smallest found so far. By considering the inner diameter of the AFI channels (0.73 ± 0.01 nm), and taking the distance between graphite sheets (0.34 nm) as the separation between the nanotube carbon atoms and the oxygen atoms on the channel wall, the diameter of the SWNTs allowed in the channels is 0.39 ± 0.01 nm (0.73 ± 0.01 nm – 0.34 nm).

There are three possible nanotube structures with a diameter of about 0.4 nm: these are the zigzag (5,0) (diameter, $d = 0.393$ nm), the armchair (3,3) ($d = 0.407$ nm) and

the chiral (4,2) ($d = 0.414$ nm), where (m, n) is a roll-up vector defining the tube symmetry. Of these, the zigzag (5,0) tube can be exactly capped by half a C_{20} fullerene⁹, which makes it more likely that this structure is formed. Because of the strong bending of carbon bonds, plus the size-confinement effect of the channels and the imperfection of such ultrasmall SWNTs (for example, point defects), there could be high-energy regions in these nanotubes. When they are released from their channels, the defects may cause instability which then permeates the whole tube, leading to the formation of the observed graphite stripes.

Because our SWNTs are highly aligned and uniform in size, their electrical transport properties are easy to measure. Local density functional calculations indicate that when the diameter of the SWNT is smaller than 0.5 nm, strong curvature effects induce strong σ – π mixing of the unoccupied orbitals, so metallicity can no longer be predicted by the simple band-folding picture¹⁰. Such small-radius nanotubes generally have finite density of states at the Fermi level. Preliminary electrical measurements on AFI crystals containing SWNTs¹¹ indicate that the temperature dependence of the SWNT conductivity is consistent with variable-range hopping in a disordered one-dimensional metallic system.

N. Wang, Z. K. Tang, G. D. Li, J. S. Chen*

Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, PR China

e-mail: phzktang@ust.hk

*Present address: Department of Chemistry, Jilin University, PR China

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Tang, Z. K., Sun, H. D., Wang, J. S., Chen, J. S. & Li, G. D. *Appl. Phys. Lett.* **73**, 2287–2289 (1998).
3. Meier, W. M. & Olson, D. H. (eds) *Atlas of Zeolite Structure Types* (Butterworth-Heinemann, London, 1992).
4. Sun, H. D., Tang, Z. K., Chen, J. S. & Li, G. D. *Appl. Phys. A* **69**, 381–384 (1999).
5. Rao, A. M. et al. *Phys. Rev. Lett.* **84**, 1820–1823 (2000).
6. Sun, L. F. et al. *Nature* **403**, 384–384 (2000).
7. Piskoti, C., Yarger, J. & Zettl, A. *Nature* **393**, 771–774 (1998).
8. Ajayan, P. M. & Iijima, S. *Nature* **358**, 23 (1992).
9. Prinzbach, H. et al. *Nature* **407**, 60–63 (2000).
10. Blau, X., Benedict, L. X., Shirley, E. I. & Louie, S. G. *Phys. Rev. Lett.* **72**, 1878–1881 (1994).
11. Tang, Z. K., Sun, H. D. & Wang, J. *Physica B* **279**, 200–203 (2000).

Neurobiology

Hydrodynamic stimuli and the fish lateral line

Sensory systems need to distinguish biologically relevant stimuli from background noise. Here we investigate how the lateral-line mechanosensory system of the fish senses minute water motions¹ in the vicinity while exposed to running water. We find that one class of receptor in

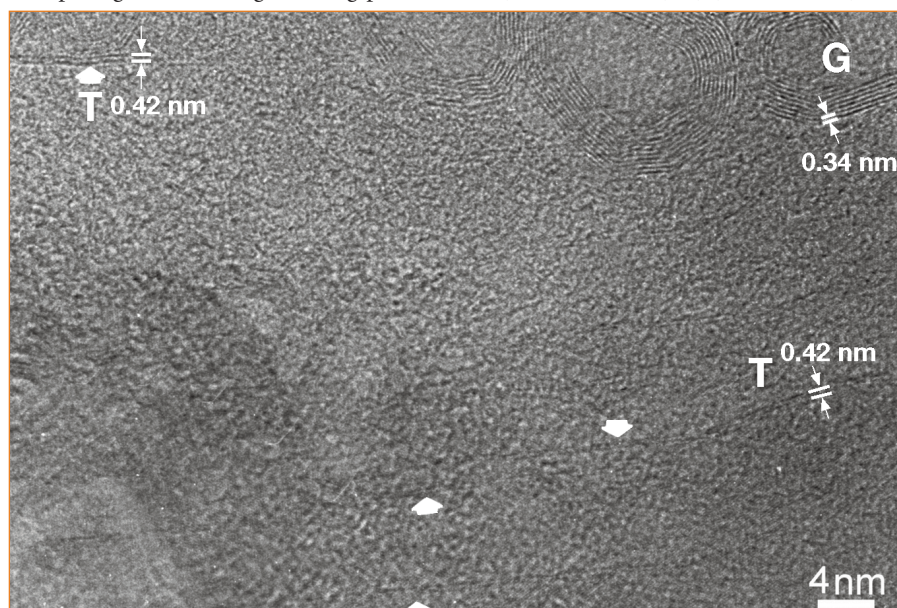


Figure 2 High-resolution transmission electron microscope image showing single-walled carbon nanotubes (indicated by T and large arrowheads) coexisting with graphite in raft-like stripes (G). More than ten single-walled carbon nanotubes are visible.

the lateral line, the canal neuromasts, can respond to hydrodynamic stimuli even in the presence of unidirectional water flow, whereas superficial neuromasts, which predominate in still-water fish, cannot.

The fish lateral line is a sensory system that contains organs known as superficial neuromasts on the skin and canal neuromasts recessed in fluid-filled subepidermal canals². Superficial neuromasts detect the velocity of water flow resulting from, for example, the fishes' own movements or from background currents, whereas canal neuromasts, because of the mechanical filter properties of lateral-line canals, are sensitive to water acceleration^{3,4}. We have investigated the effects of these separate functions of the lateral-line periphery on the detection of stimuli presented in the presence of running water.

We used the still-water fish *Carassius auratus* ($n=18$) to measure the neural responses of primary afferents in the posterior lateral-line nerve to the water motions caused by a stationary vibrating sphere in still and in running water. In still water, we distinguished two types of afferent: type I afferents ($n=15$) responded in proportion to the velocity of water motions and must therefore innervate superficial neuromasts³; type II afferents ($n=11$) responded roughly in proportion to the acceleration component of water motions and must therefore innervate canal neuromasts³. Temporal response profiles and the degree of phase coupling to the stimulus were comparable for both types of afferent (U -test, $P>0.1$).

Type I afferents responded to unidirectional water flow (10 or 15 cm s^{-1}) with a burst-like increase in discharge rate (Wilcoxon test, $P\leq 0.011$). This was most probably due to microturbulence close to the fish, as measured by particle-imaging velocimetry⁵. Discharge rates of type II afferents did not increase in running water (Wilcoxon test, $P\geq 0.067$). The responses of type I afferents to sinusoidal water motions were masked in the presence of unidirectional water flow (Wilcoxon test, $P\leq 0.041$), whereas the responses of type II afferents were not (Wilcoxon test, $P\geq 0.12$) (Fig. 1a).

At stimulus displacement amplitudes greater than 20 μm , all afferents phase-coupled to the vibrating sphere, whereas phase coupling of type I afferents was significantly reduced in running water (Wilcoxon test, $P\leq 0.036$) (Fig. 1b). In contrast, phase coupling of type II afferents at 10 cm s^{-1} and 15 cm s^{-1} flow velocity did not differ from phase coupling in still water (Wilcoxon test, $P\geq 0.11$).

Our results distinguish the function of the two classes of lateral-line receptor. Superficial neuromasts (innervated by type I afferents) are highly sensitive to a vibrating sphere only in still water. In contrast,

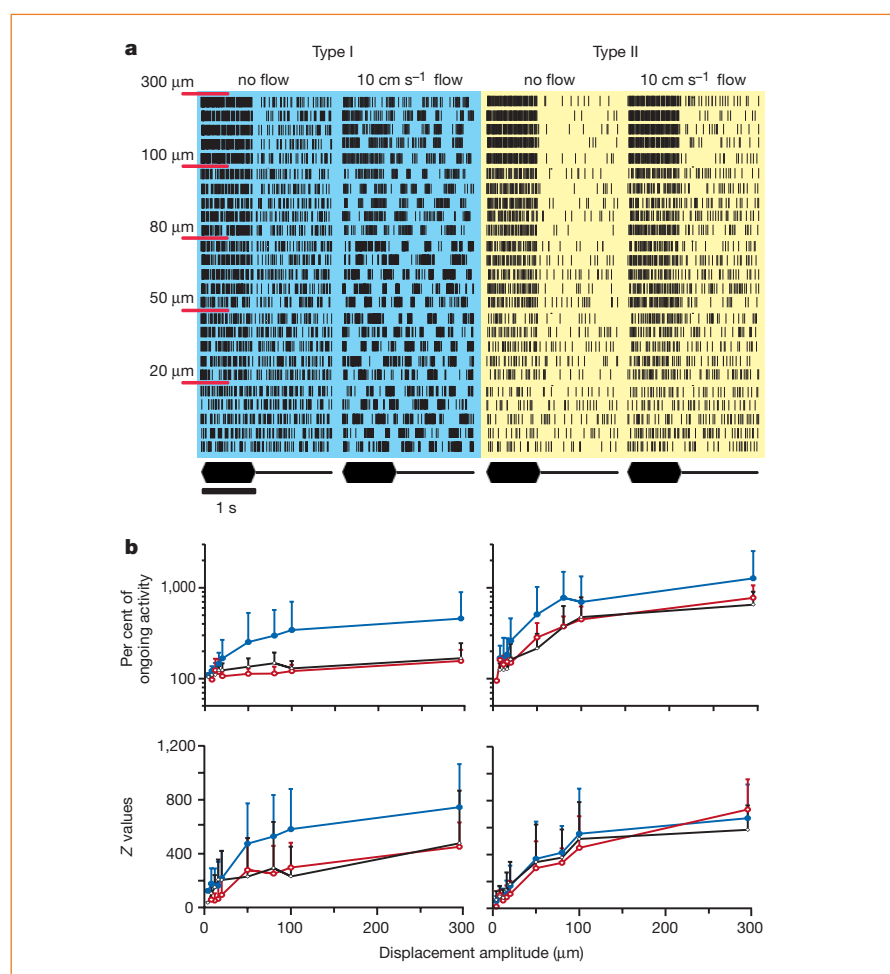


Figure 1 The effect of unidirectional water flow on the responses of primary lateral-line afferents to a stationary vibrating sphere (diameter 10 mm). Fish were positioned towards the flow. **a**, Raster diagram of the responses of a type I (blue background) and a type II (yellow background) afferent. Responses to five stimulus-repetitions are shown. Peak-to-peak sphere displacement amplitudes varied between 300 μm and 20 μm ; vibration frequency was 50 Hz. The stimulus trace is shown underneath. **b**, Percentage of ongoing activity and Z values as a measure of phase coupling¹¹ of type I and type II afferents as a function of sphere displacement for still water (blue) and running water (red: 10 cm s^{-1} ; black: 15 cm s^{-1}) conditions. The percentage of ongoing activity was calculated relative to the discharge rate in still and running water, respectively, but without sphere vibration and was averaged across fibres. Vertical bars represent one standard deviation.

unidirectional water flow hardly affects the responses of canal neuromasts (type II afferents). Fishes that live in running or turbulent water or are fast swimmers have a well-developed canal system^{6,7}, whereas closely related species that live in still waters and are slow swimmers or are sedentary have many superficial neuromasts but a less well-developed canal system^{8,9}.

In running water and/or in swimming fish, superficial neuromasts are permanently stimulated. This makes them useful for rheotaxis¹⁰ but useless for the detection of small hydrodynamic stimuli against background noise. Canal neuromasts, however, are efficient at detecting hydrodynamic stimuli even in the constant presence of background water flow, because of the filter properties of lateral-line canals. This may help alert the fish to vibrations made by nearby prey or an approaching predator.

J. Engelmann, W. Hanke, J. Mogdans, H. Bleckmann

Institut für Zoologie der Rheinischen Friedrich-Wilhelms-Universität Bonn, Poppelsdorfer Schloss, 53115 Bonn, Germany
e-mail: bleckmann@uni-bonn.de

- Bleckmann, H. *Progress in Zoology* Vol. 41 (ed. Rathmayer, W.) 1–115 (Fischer, Stuttgart, 1994).
- Northcutt, R. G. in *The Mechanosensory Lateral Line: Neurobiology and Evolution* (eds Coombs, S., Görner, P. & Münz, H.) 17–78 (Springer, New York, 1989).
- Kroese, A. B. A. & Schellart, N. A. M. *J. Neurophysiol.* **68**, 2212–2221 (1992).
- Denton, E. J. & Gray, J. A. B. in *The Mechanosensory Lateral Line: Neurobiology and Evolution* (eds Coombs, S., Görner, P. & Münz, H.) 229–246 (Springer, New York, 1989).
- Hanke, W., Brückner, C. & Bleckmann, H. *J. Exp. Biol.* **203**, 1193–1200 (2000).
- Jakubowski, M. *Acta Biol. Cracov. Ser. Zool.* **10**, 69–81 (1967).
- Bleckmann, H. & Münz, H. *Brain Behav. Evol.* **35**, 240–250 (1990).
- Dijkgraaf, S. *Biol. Rev.* **38**, 51–106 (1963).
- Marshall, N. B. *Explorations in the Life of Fishes* (Harvard Univ. Press, Cambridge, MA, 1971).
- Montgomery, J. C., Baker, C. F. & Carton, A. G. *Nature* **389**, 960–963 (1997).
- Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1981).

Telomere states and cell fates

Elizabeth H. Blackburn

Departments of Biochemistry and Microbiology, University of California, San Francisco, San Francisco, California 94143-0448, USA

Telomere length has frequently been used as a means to predict the future life of cells. But by itself it can be a poor indicator of ageing or cell viability. What, then, is the important property of a telomere? Here recent findings are integrated into a new, probabilistic view of the telomere to explain how and when it can signal not only its own fate but also that of a cell.

Telomeres are special functional complexes at the ends of eukaryotic chromosomes. Linear DNA molecules, such as eukaryotic chromosomal DNAs, require mechanisms besides the conventional DNA polymerases to complete the replication of their ends. This is achieved by the ribonucleoprotein enzyme telomerase, which balances terminal DNA losses by lengthening the ends of eukaryotic telomeric DNA through RNA-templated addition of tandemly repeated telomeric sequences. Another hazard of having the chromosomal DNA in linear form is that telomeres must be prevented from undergoing end-to-end fusion, which is the normal way of repairing chromosome-internal DNA ends produced by DNA breakage. Telomere fusion results in chromosome instability and must therefore be avoided.

In proliferating cells lacking functional telomerase, the population of heterogeneously sized telomeres progressively shortens and the cell population as a whole eventually undergoes senescence. Experimentally disrupting telomerase action causes telomere shortening and cellular senescence¹ (also reviewed in ref. 2). Conversely, in some experiments activating telomerase in certain human somatic cells, which lack telomerase activity when grown in primary culture, caused telomere lengthening and immortalized the cells^{3,4}. Hence the view emerged that a 'critical length' was the property of telomeres that caused cellular senescence.

Recent work in several systems has uncovered a more complex relationship between telomere length and cell proliferative capacity. Here a new way to link telomeres to cellular senescence is proposed. The central concept is that the telomere is a dynamic nucleoprotein complex that can switch stochastically between two states: capped and uncapped. Capping is functionally defined as preserving the physical integrity of the telomere, allowing cell division to proceed. However, regulated uncapping occurs normally in dividing cells, with the crucial property of a functional telomere being a high probability that it will rapidly switch back to a capped state. Left uncorrected too long, the uncapped state elicits cell-cycle arrest or other responses. Crucially, this probabilistic model identifies telomere capping status, as opposed to telomere length alone, as the important property relevant to telomere function. Factors, such as telomere length, that impinge on telomere function can now be evaluated in terms of their effects on telomere capping states.

Telomere length does not seal the fate of a cell

Telomere shortness has often been assumed to indicate the number of times a cell has divided and can be expected to divide. For cells in culture lacking telomerase, this mitotic 'clock' theory holds true experimentally^{5,6}. It is also applicable to knockout mice lacking active telomerase, which display progressive telomere shortening, both as the animals age and over increasing generations^{7,8}. In some normal somatic cells of adult humans and wild mice telomere lengths decrease with age^{9,10}, leading to the suggestion that telomere shortening contributes to the organismal ageing by limiting cell proliferation. Recent studies show, however, that cultured fibroblasts (which lack active telomerase) from older people reached

senescence *in vitro* no faster than fibroblasts taken from younger people, suggesting that limited cell proliferation *in vitro* is not simply related to age in humans¹¹. Furthermore, when telomerase is active a 'mitotic clock' model does not apply; by stabilizing telomeres, telomerase effectively moves back the hands of any such 'clock'. This is of direct interest for human studies because telomerase is active not only in human germline, embryonic and stem cells, but also in several types of normal adult proliferating somatic cells—in the immune system, skin, intestinal lining and hair follicles. Telomeres in human blood cell populations show complex length changes with age¹², actually lengthening in certain immune system cells of adult humans and toads^{13,14}.

Telomere length is strongly influenced by many genetic, and other factors. Thus, in the same cell type telomeres can be longer in an old person than in another, younger person^{10,15}. Hence only when genetic and other conditions are strictly controlled are the factors determining telomere functionality unmasked.

One such factor is the presence of active telomerase itself. In two budding yeasts, *Saccharomyces cerevisiae* and *Kluyveromyces lactis*, cells lacking catalytically active telomerase ceased dividing after a period of telomeric shortening. However, cells of otherwise isogenic strains expressing certain hypomorphic, but catalytically fully active, mutant telomerases continued to divide normally and indefinitely, despite their telomeres becoming and remaining much shorter than the shortest telomeres in the telomerase-lacking control cells, which had stopped dividing. This was found with a variety of hypomorphic telomerases that retained enzyme activity but caused telomeres to be short but stable^{16,17} (also reviewed in ref. 18). Telomere shortness may therefore be clearly uncoupled from loss of function, with the presence or absence of active telomerase being the critical factor.

Strikingly similar results have been obtained in human cells. Normal cultured endothelial cells and SV40 T-antigen-transformed fibroblasts lack telomerase activity, and telomeres shorten as they divide. Without telomerase activation, the endothelial cells senesce and the transformed human fibroblasts undergo crisis, characterized by chromosomal instabilities and failure to proliferate. Expressing telomerase from a constitutive retroviral promoter effectively immortalized both cell types. As in the yeast experiments, although the cells continued to proliferate, the telomeres shortened further and remained stably shorter than in the control cells at the point when they had stopped proliferation^{19,20}. Similar results were found with transformed fibroblasts that had spontaneously immortalized and activated their endogenous telomerase²¹. Remarkably, the transformed fibroblasts with experimentally activated telomerase also had fewer chromosomal end-to-end fusions (indicative of telomere failure) than did control cells, which entered crisis despite their longer telomeres. Hence, in yeasts and human cells, enzymatically competent telomerase has a protective effect on very short telomeres that, in its absence, would have caused cells to stop dividing. Notably, the presence of enzymatically competent telomerase is not apparently required when telomeres are sufficiently long; only when telomere length falls into a certain range does the

presence or absence of telomerase become a critical factor^{19,22}.

A two-state model of telomere capping

If telomere length does not always indicate whether a telomere is functional, what does? As telomeres protect the ends of chromosomes from fusion (end-to-end joining) and its deleterious consequences, telomere capping has been described as protecting the telomere from being seen as a broken end. A DNA break within a chromosome is sensed and responded to by the DNA damage-response machinery of the cell, with the result that, following cell-cycle arrest to allow time for the repair, the DNA break is fixed by end-to-end rejoining. Yet surprisingly, the DNA end-joining response protein Ku is specifically located at telomeres^{23,24}. Moreover, Ku, as well as certain components of other DNA damage-response pathways, are actually needed for telomere maintenance^{25–28}. Mutating the genes encoding DNA damage response kinases (*TEL1* and *MEC1* in budding yeast, *tel1+* and *rad3+* in fission yeast, and ATM in humans), a DNA damage checkpoint protein (MRT-2 in the nematode), or poly-(ADP ribose) polymerase, which acts at DNA breaks, causes telomeres to become short and prone to fusions^{28–30} (also reviewed in ref. 18). Importantly, genetic experiments show that *TEL1* and *MEC1* (and probably MRT-2 in the nematode) are needed specifically for telomerase to act on telomeres²⁷. Therefore, the telomere actually has to look like a DNA break for telomerase to act and thereby recap the telomere. This implies that a telomere continually sustains some judiciously regulated degree of uncapping. These findings turn on its head the previous idea of how a telomere protects a chromosome end. They can be integrated with the molecular information on telomeres into a dynamic two-state model for telomeres, involving their switching between uncapped or capped states (Fig. 1).

First, this model provides a new framework to rationalize the recent results linking telomeres and DNA damage signalling in cells. The DNA damage response elicited on sensing an uncapped telomere or other broken DNA end brings important enzymes to these ends. Now telomere function can be envisaged as regulating and channelling the active and sensitive surveillance DNA damage response, which can detect even a single DNA break in a cell, into an appropriate telomere-specific response that maintains the integrity

of the telomere. Thus, a crucial telomere-specific property of an uncapped telomere is to divert the DNA damage response into telomerase action rather than into DNA end-to-end joining³¹. As described above, active telomerase helps even very short telomeres to be functionally capped. Possibly, once targeted to the telomere by the DNA damage response, the active telomerase in turn signals back to the cell, reassuring it to keep dividing. Another appropriate response to an uncapped telomere is homologous recombination, which in the absence of telomerase commonly acts on telomeric DNA regions to extend them²². Only when all these processes fail do cell-cycle exit or telomere end-joining ensue.

Second, the model ties molecular properties of a telomere, and telomere length, to capping. Much evidence fits a model in which the high local concentration of proteins nucleated on the tandem-telomeric repeat array is conducive to formation of a highly cooperative complex (see Box 1). In this model, the complex can switch between physical states corresponding to functionally capped and uncapped. Longer telomeres (with more repeats and thus more protein-binding sites) are more likely to switch into the capped state structure than are shorter telomeres^{32,33}. The model can explain why, in cells with telomerase, the lengths of the population of telomeres are kept within well-defined limits. A shortened telomere elicits telomerase action, increasing its probability of switching to the capped state. On recapping, the resulting telomeric DNA–protein complex acts as a gatekeeper to control telomerase action (as well as recombination and degradative activities associated with repair pathways). Shortening by incomplete replication or degradation increases the probability of uncapping again. In addition to length, active telomerase and several other molecular components of the telomeric DNA–protein complex collaborate synergistically to affect this dynamic balance between capped and uncapped states^{33–36} (see Box 1).

Two states of cells

Cellular senescence is commonly described as cells being able to divide a finite number of times, with senescence occurring only when telomeres reach a critical short length. Usually implicit is the idea that the ‘young’ cells, which are all dividing at the beginning of the passing, are free of ‘aged’ phenotypes until, following a period

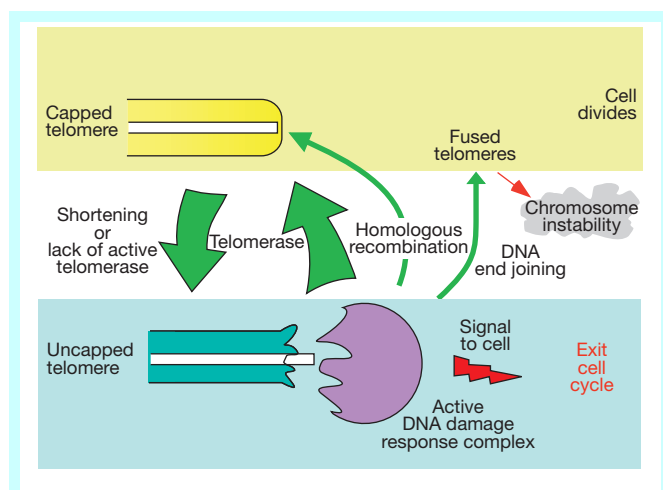


Figure 1 Telomere switching. The appropriate response to the uncapping of a telomere is action by telomerase (primarily) or homologous recombination, protecting and/or elongating the telomere so that cell cycling can resume. Non-homologous end-joining of telomeres can occur, fusing them and removing the immediate damage signal, but when cell divisions resume the fused chromosomes are unstable. If none of these ways of capping occurs, the response of a normal cell is exit from the cell cycle or, in certain mammalian cells, cell suicide (apoptosis)^{47,48}.

Box 1

Capping function is provided by a mutually reinforcing set of molecular components

The tandem array of telomeric-DNA repeats at each chromosome end attracts and binds a set of sequence-specific DNA-binding proteins. These in turn bind a further set of proteins to assemble an inferred higher order complex nucleated on the telomeric DNA repeats^{18,33,44}. The telomeric complex exists in two states that determine telomere capping. These may be analogous to the repressed (silenced) and non-repressed states of heterochromatin, as some of the proteins in the telomeric complex that affect capping in yeast are the same ones that silence expression of genes placed near a telomere or at the mating type loci (reviewed in ref. 33). Such silencing is characterized by epigenetic states that are stable for several cell divisions. Molecular evidence for similarly semi-stable telomere capping is seen experimentally in telomeres that have had one or more components mutated: they show sudden changes in telomere length regulation, switching abruptly and stochastically to an unregulated state^{32,45}.

Known molecular determinants of capping include active telomerase^{19,45}, structural proteins of the higher order telomeric complex, and particularly, the properties of the DNA–protein complexes at the very tip of the telomeric tract^{31,35,36,45–47}. All these components act synergistically in capping: compromising one can be harmless, but disrupting two or more often causes uncapping^{35,46}. Thus, whether a telomere of a given length is uncapped depends on the status of the other capping components of a telomere.

of telomere shortening during the passaging, critically short telomeres appear in the 'old' cells. Implicit is the idea of an inbuilt delay before senescence is reached. But a significant (although infrequently cited) literature dealing with the behaviour of individual cells reveals a different picture (for example, refs 37, 38; reviewed in ref. 39).

Quantitative examinations, on a cell-by-cell basis, of normal human and mouse somatic cells in culture have shown that these cells exist in two states: cycling and exited from the cell cycle. Even from the very beginning of passaging, cells are stochastically dropping out of the cycling population. They do so with ever-increasing frequency until the population as a whole ceases doubling³⁸. Notably, even late in the passaging any cells still in the cycling state are cycling as fast as those early in the passaging^{38,40}. Hence, the only difference between early and late passages is the frequency of cycling versus non-cycling cells; those cycling cells that are 'old' by virtue of their passage number are not significantly different from the cycling cells early in the passaging.

This well-documented stochastic two-state behaviour of individual cells in a population matches remarkably well with the behaviour predicted if the shortening telomeres in cells dividing without telomerase are in a two-state population as described above (Fig. 2). In this model, even at the beginning of the passaging, the telomeres, although relatively long, are predicted to have a finite (although initially low) probability of becoming uncapped, as there is no telomerase present. The lack of telomerase precludes recapping, and the uncapped telomere signals the cell to exit the cell cycle. As the cells undergo their divisions without telomerase, the telomere population as a whole shortens, increasing the probability that at any point a telomere will stochastically switch to an uncapped state. The probability of any cell exiting the cell cycle correspondingly increases during culturing.

This new framing of the connection between telomere shortening and cellular senescence differs from former descriptions by introducing the concept of a stochastic and increasing probability of

switching to the uncapped/non-cycling state. It can explain several observations. In human cells in culture, two mitotic sister cells can have vastly different proliferative potentials³⁷ even though they have telomeres that are similar in length (reviewed in ref. 39). Simple stochastic switching behaviour of telomeres accounts for both the growth kinetics of cultured cells, discussed above, and the large amount of information about the molecular components of telomere capping. It can also account for the observation that in telomerase knockout mice, the telomere fusions and phenotypic effects of telomerase deficiency increase steadily in frequency and severity with increasing generations, rather than showing up abruptly only in late generations^{7,41,42}. In yeasts, cellular senescence is also stochastic and progressive in the cell population (see, for example, ref. 22). Importantly, the new model obviates the need to define a 'critical' telomere length. Instead, whether a telomere will become uncapped is expressed as a probabilistic function influenced by several factors, only one of which is its length.

Implications

Understanding how telomeres function and how telomerase exerts its protective effects will be necessary if telomerase is to be applied effectively for medical use. New results have necessitated the rethinking of some previously held ideas about telomeres. The accumulated data show that telomere length alone cannot be taken as an indicator of the number of times the cells can or did divide, nor of the age of an organism. Thus, we have to look beyond length to understand telomere functionality. New questions have arisen: what is the precise molecular nature of the two telomeric states? How do the many identified parts of this cellular systems—telomerase, the DNA damage-response pathways and the different components of telomeres—act together to ensure telomeric and hence genomic stability? Such interactions can impact on the proliferation of both healthy and cancerous cells. Increasingly, investigations into a variety of biological processes point to the view that stochastic stabilization of one of two functionally different physical states can underlie many seemingly deterministic phenomena in biology⁴³. The stochastic behaviours of telomeres and cells may provide another example of this. □

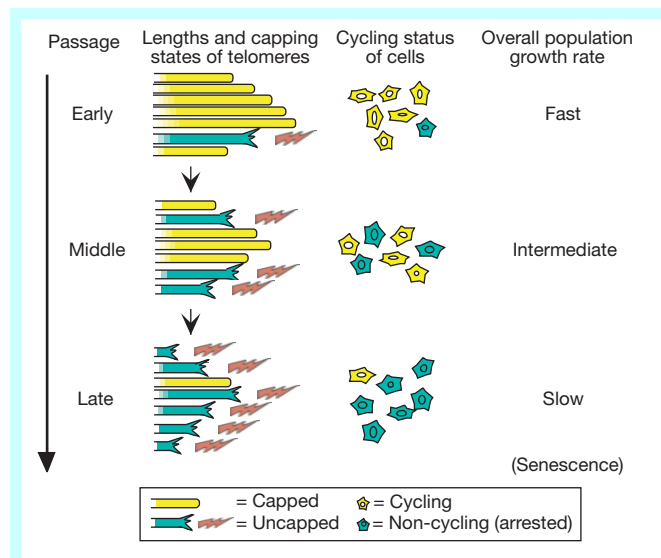


Figure 2 Stochastic uncapping of telomeres and cellular senescence kinetics. A probabilistic two-state switching behaviour of telomeres accounts for the two-state cell kinetics of primary cultures of normal mouse and human glial cells and fibroblasts^{38,40}. Early in culturing most telomeres are functionally capped (yellow) and most cells are cycling (yellow cells), but, lacking telomerase, a finite fraction of the telomeres stochastically became uncapped (blue), causing cell-cycle exit (blue cells). This fraction increases steadily during culturing as the telomeres shorten, increasing their probability of stochastically uncapping, until net growth of the cell population ceases (senescence).

- Yu, G. L., Bradley, J. D., Attardi, L. D. & Blackburn, E. H. *In vivo* alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs. *Nature* **344**, 126–132 (1990).
- Nugent, C. I. & Lundblad, V. The telomerase reverse transcriptase: components and regulation. *Genes Dev.* **12**, 1073–1085 (1998).
- Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
- Counter, C. M. *et al.* Dissociation among *in vitro* telomerase activity, telomere maintenance, and cellular immortalization. *Proc. Natl Acad. Sci. USA* **95**, 14723–14728 (1998).
- Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
- Counter, C. M. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
- Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
- Prowse, K. R. & Greider, C. W. Developmental and tissue specific regulation of mouse telomerase and telomere length. *Proc. Natl Acad. Sci. USA* **92**, 4818–4822 (1995).
- Harley, C. B. Human ageing and telomeres. *Ciba Foundation Symp.* **211**, 129–139 (1997).
- Cristofalo, V. J., Allen, R. G., Pignolo, R. J., Martin, B. G. & Beck, J. C. Relationship between donor age and the replicative lifespan of human cells in culture: a reevaluation. *Proc. Natl Acad. Sci. USA* **95**, 10614–10619 (1998).
- French, R. W. Jr, Blackburn, E. H. & Shannon, K. M. The rate of telomere sequence loss in human leukocytes varies with age. *Proc. Natl Acad. Sci. USA* **95**, 5607–5610 (1998).
- Weng, N. P., Hathcock, K. S. & Hodes, R. J. Regulation of telomere length and telomerase in T and B cells: a mechanism for maintaining replicative potential. *Immunity* **9**, 151–157 (1998).
- Bassham, S., Beam, A. & Shampay, J. Telomere variation in *Xenopus laevis*. *Mol. Cell. Biol.* **18**, 269–275 (1998).
- Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).
- Prescott, J. & Blackburn, E. H. Telomerase RNA mutations in *Saccharomyces cerevisiae* alter telomerase action and reveal nonprocessivity *in vivo* and *in vitro*. *Genes Dev.* **11**, 528–540 (1997).
- Roy, J., Fulton, T. B. & Blackburn, E. H. Specific telomerase RNA residues distant from the template are essential for telomerase function. *Genes Dev.* **12**, 3286–3300 (1998).
- Blackburn, E. H. *et al.* in *Cold Spring Harbor Symposia on Quantitative Biology: Biological Responses to DNA Damage* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, in the press).

19. Zhu, J., Wang, H., Bishop, J. M. & Blackburn, E. H. Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. *Proc. Natl Acad. Sci. USA* **96**, 3723–3728 (1999).
20. Yang, J. *et al.* Human endothelial cell life extension by telomerase expression. *J. Biol. Chem.* **274**, 26141–26148 (1999).
21. Ducray, C., Pommier, J. P., Martins, L., Boussin, F. D. & Sabatier, L. Telomere dynamics, end-to-end fusions and telomerase activation during the human fibroblast immortalization process. *Oncogene* **18**, 4211–4223 (1999).
22. McEachern, M. J. & Blackburn, E. H. Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes Dev.* **10**, 1822–1834 (1996).
23. Gravel, S., Larrivée, M., Labrecque, P. & Wellinger, R. J. Yeast Ku as a regulator of chromosomal DNA end structure. *Science* **280**, 741–744 (1998).
24. Hsu, H. L., Gilley, D., Blackburn, E. H. & Chen, D. J. Ku is associated with the telomere in mammals. *Proc. Natl Acad. Sci. USA* **96**, 12454–12458 (1999).
25. Bertuch, A. & Lundblad, V. Telomeres and double-strand breaks: trying to make ends meet. *Trends Cell Biol.* **8**, 339–342 (1998).
26. Le, S., Moore, J. K., Haber, J. E. & Greider, C. W. RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* **152**, 143–152 (1999).
27. Ritchie, K. B., Mallory, J. C. & Petes, T. D. Interactions of TLC1 (which encodes the RNA subunit of telomerase), TEL1, and MEC1 in regulating telomere length in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6065–6075 (1999).
28. Ahmed, S. & Hodgkin, J. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**, 159–164 (2000).
29. Sprung, C. N., Bryan T. M., Reddel, R. R. & Murnane, J. P. Normal telomere maintenance in immortal ataxia telangiectasia cell lines. *Mutat. Res.* **379**, 177–184 (1997).
30. d'Adda di Fagagna, F. *et al.* Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. *Nature Genet.* **23**, 76–80 (1999).
31. Evans, S. K. & Lundblad, V. Est1 and Cdc13 as comediators of telomerase access. *Science* **286**, 117–120 (1999).
32. McEachern, M. J. & Blackburn, E. H. Runaway telomere elongation caused by telomerase RNA gene mutations. *Nature* **376**, 403–409 (1995).
33. Marcand, S., Gilson, E. & Shore, D. A protein-counting mechanism for telomere length regulation in yeast. *Science* **275**, 986–990 (1997).
34. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
35. Krauskopf, A. & Blackburn, E. H. Rap1 protein regulates telomere turnover in yeast. *Proc. Natl Acad. Sci. USA* **95**, 12486–12491 (1998).
36. Smith, C. D. & Blackburn, E. H. Uncapping and deregulation of telomeres lead to detrimental cellular consequences in yeast. *J. Cell Biol.* **145**, 203–214 (1999).
37. Jones, R. B., Whitney, R. G. & Smith, J. R. Intramitotic variation in proliferative potential: stochastic events in cellular aging. *Mech. Age. Dev.* **29**, 143–149 (1985).
38. Pontana, J., Stein, W. D. & Shall, S. A quantitative analysis of the aging of human glial cells in culture. *J. Cell. Phys.* **117**, 342–352 (1983).
39. Holliday, R. Endless quest. *BioEssays* **18**, 3–5 (1983).
40. Karatza, C., Stein, W. D. & Shall, S. Kinetics of *in vitro* ageing of mouse embryo fibroblasts. *J. Cell Sci.* **65**, 163–175 (1984).
41. Herrera, E., Samper, E. & Blasco, M. A. Telomere shortening in mTR^{-/-} embryos is associated with failure to close the neural tube. *EMBO J.* **18**, 1172–1181 (1999).
42. Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
43. Kirschner, M., Gerhart, J. & Mitchison, T. Molecular “Vitalism”. *Cell* **100**, 79–88 (2000).
44. Li, B., Oestreich, S. & de Lange, T. Identification of human Rap1: Implications for telomere evolution. *Cell* **101**, 471–483 (2000).
45. Blackburn, E. H. The telomere and telomerase: how do they interact? *Mt Sinai J. Med.* **66**, 292–300 (1999).
46. Krauskopf, A. & Blackburn, E. H. Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats. *Nature* **383**, 354–357 (1996).
47. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321–1325 (1999).
48. Kondo, S. *et al.* Antisense telomerase treatment: induction of two distinct pathways, apoptosis and differentiation. *FASEB J.* **12**, 801–811 (1998).

Acknowledgements

I thank C. Gross, I. Herskowitz, B. Panning and C. Smith for discussions and comments on the manuscript.

Correspondence should be addressed to E.H.B. (e-mail: telomer@itsa.ucsf.edu).

Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function

Julia Parrish-Novak*, Stacey R. Dillon†, Andrew Nelson†, Angie Hammond*, Cindy Sprecher*, Jane A. Gross†, Janet Johnston†, Karen Madden*, Wenfeng Xu*, Jim West‡, Sara Schrader*, Steve Burkhead*, Mark Heipel*, Cameron Brandt*, Joseph L. Kuijper*, Janet Kramer*, Darrell Conklin§, Scott R. Presnell§, Jon Berry‡, Faith Shiota†, Susan Bort†, Kevin Hambly†, Sherri Mudri†, Chris Clegg†, Margaret Moore||, Francis J. Grant§, Catherine Lofton-Day||, Teresa Gilbert||, Fenella Raymond||, Andrew Ching§, Lena Yao||, Deb Smith||, Philippa Webster||, Theodore Whitmore||, Mark Maurer||, Kenneth Kaushansky¶, Rick D. Holly* & Don Foster*

Departments of * Functional Cloning, † Immunology, ‡ Protein Biochemistry, § Biomolecular Informatics, and || Genetics, ZymoGenetics, Inc., 1201 Eastlake Avenue E., Seattle, Washington 98102, USA

¶ Division of Hematology, Box 35 7710, University of Washington, School of Medicine, 1959 NE Pacific Street, Seattle, Washington 98195, USA

Cytokines are important in the regulation of haematopoiesis and immune responses, and can influence lymphocyte development. Here we have identified a class I cytokine receptor that is selectively expressed in lymphoid tissues and is capable of signal transduction. The full-length receptor was expressed in BaF3 cells, which created a functional assay for ligand detection and cloning. Conditioned media from activated human CD3⁺ T cells supported proliferation of the assay cell line. We constructed a complementary DNA expression library from activated human CD3⁺ T cells, and identified a cytokine with a four-helix-bundle structure using functional cloning. This cytokine is most closely related to IL2 and IL15, and has been designated IL21 with the receptor designated IL21R. *In vitro* assays suggest that IL21 has a role in the proliferation and maturation of natural killer (NK) cell populations from bone marrow, in the proliferation of mature B-cell populations co-stimulated with anti-CD40, and in the proliferation of T cells co-stimulated with anti-CD3.

Although primary sequence homology among the interleukins (ILs) and haematopoietic growth factors is low, these cytokines are structurally related¹ and fold into a classic 'four-helix-bundle' structure that is representative of the family. Most cytokines bind to cells and transduce signals through either the class I or the class II cytokine receptors. The class II cytokine receptors include the receptors for IL10 and the interferons, whereas the class I cytokine receptor family includes the receptors for IL2, IL3, IL4, IL5, IL6, IL7, IL9, IL11, IL12, IL13 and IL15, as well as the haematopoietic growth factors, leptin and growth hormone². Class I cytokine receptors that are capable of signal transduction through homodimerization have been used in the past to design biological activity assays to screen for sources of the cognate ligand, and to facilitate either purification or expression cloning of the ligand³.

Here we describe a new class I cytokine receptor, IL21R. We describe the use of a proliferation assay, based on BaF3 cells⁴ expressing IL21R, to screen for potential interactions with known cytokines, to identify a source of the cognate ligand, and to clone the ligand, IL21. Preliminary biological characterization indicates that IL21 may be involved in haematopoietic development and differentiation of natural killer cells in conjunction with IL15, in the activation and proliferation of B cells co-stimulated through CD40 and also in the co-stimulation of T cells.

Receptor cloning and assay development

We identified an expressed sequence tag (EST) containing a predicted signal peptide and a predicted amphipathic helix. Sequencing of a full-length clone revealed a cDNA coding for a member of the class I cytokine receptor family. This cDNA, designated *IL21R*, contains an open reading frame encoding a 538-amino-acid protein that has structural motifs common to class I cytokine receptors^{2,5}. Extracellular motifs include a single cytokine recognition module, two pairs of conserved cysteine residues, and a 'WSXWS' motif. The intracellular domain contains strong intracellular signalling motifs, including classical box 1 and box 2 motifs^{6–10}, which indicate that

this receptor is probably a signalling subunit. IL21R has highest amino-acid sequence similarity to IL2R β and IL4R α . We subsequently cloned murine *IL21R* from a mouse splenocyte library, and found that it shares overall structure and functional motifs with human IL21R (Fig. 1a).

IL21R was localized to human 16p11 using the GeneBridge 4 radiation hybrid mapping panel (Research Genetics, Inc.). In addition, the *IL21R* sequence was found in genomic sequence from two overlapping bacterial artificial chromosome clones mapped to 16p11 (GenBank accession nos AC002303 and AC004525). Comparison of the *IL21R* cDNA sequence with this genomic sequence revealed that the *IL21R* gene spans about 20 kilobases (kb) of genomic DNA, and lies about 65 kb from the *IL4R α* gene.

Mpl, the receptor for thrombopoietin (TPO), delivers a proliferative signal on homodimerization in the presence of TPO⁷. To determine whether the intracellular domain of IL21R is capable of signalling as a homodimer, we constructed a chimaeric receptor using the extracellular and transmembrane domains of Mpl fused to the intracellular domain of IL21R. This chimaeric receptor was transfected into BaF3 cells⁴, which are IL3-dependent and not normally responsive to TPO. Positive control cells were transfected with full-length Mpl. Both the full-length Mpl and the chimaeric receptor transduced proliferative signals in the BaF3 cells in response to TPO (data not shown). The proliferative signal delivered by the homodimeric Mpl/IL21R chimaera in BaF3 cells suggests that a ligand for IL21R may exert a proliferative effect on cells expressing the full-length receptor, thus providing an activity assay for ligand cloning.

Ligand functional cloning

We expressed full-length IL21R in BaF3 cells, creating the cell line BaF3/IL21R. Known human and murine cytokines were tested for activity, and only those that also stimulated parental BaF3 cells supported growth of BaF3/IL21R cells. Conditioned media from

A cDNA library was constructed from sorted CD3⁺ human T cells that had been activated for 13 h with PMA and ionomycin. Pools of 250 clones each were transfected into baby hamster kidney (BHK) cells, and conditioned media were transferred to the BaF3/IL21R

Ligand structure

The sequence of human *IL21* cDNA contains an open reading frame that encodes a polypeptide precursor of 162 amino acids. The signal

Figure 1 Amino-acid sequences of IL21R and IL21. **a**, Alignment of the amino-acid sequences of murine and human IL21R. Overall identity is 62%. Signal sequence is underlined. Both pairs of cysteines (shown in bold) and the WSEWS motif (underlined) are fully conserved in the mouse. Box 1 and Box 2 (shown in bold) are less conserved, but conform to the overall consensus sequences. A consensus Stat3-binding site²⁸,

underlined and in boldface, is present near the C terminus. **b**, Alignment of human and murine IL21 with related cytokines. Identities (colon) or similarities (asterisk) between either human or murine IL21 and human IL15 are indicated. Mature N termini are indicated by open boxes. Potential *N*-linked glycosylation sites are underlined.

peptidase cleavage rules¹¹ predict a cleavage site after Gly 31, leaving a mature N terminus that corresponds to our observed N terminus for recombinant human IL21 protein (data not shown). The mature polypeptide has a predicted relative molecular mass of 15,000 (M_r 15K), and consists of a 131-residue four-helix-bundle cytokine domain with significant homology to IL2, IL4 and IL15 (Fig. 1b). IL21 and IL15 share two pairs of cysteine residues in identical positions, one pair that is conserved in IL2, IL4 and granulocyte-macrophage colony-stimulating factor (GM-CSF), and one that is unique to IL21 and IL15. This structural consistency, together with the relatively high degree of amino-acid homology, indicates that IL15 is the closest structural relative of IL21.

Using the Stanford G3 radiation hybrid panel (Research Genetics, Inc.), we mapped the human *IL21* gene to 4q26-q27, roughly 180 kb from the *IL2* gene. The *IL15* gene lies more distal in the cluster. The presence of these three highly related genes in the same chromosomal region suggests that they may have arisen by gene duplication.

Murine IL21

A search of nucleic-acid databases using the human *IL21* sequence revealed a partial, unspliced expressed sequence tag (EST1483966) representing the murine orthologue of *IL21*. We obtained a cDNA clone by screening a library constructed from activated CD90⁺ splenocytes. Murine *IL21* cDNA encodes a precursor with a signal peptide and a 122-amino-acid mature protein with 57% identity to human *IL21* (Fig. 1b). The sequences of human and murine *IL21* are especially well conserved in the regions predicted to encode α -helices A and D, regions that in other cytokines are important in receptor interactions^{1,12}. Both human and murine IL21 have an acidic amino acid in position Asp13 of the mature peptide, corresponding to Glu9 in IL4, which is involved in the primary high-affinity interaction between IL4 and IL4R α ¹². Human and murine IL21 also contain a conserved Gln114, equivalent to Gln141 of IL2, which has been implicated in the interaction between IL2 and IL2R γ c (ref. 13).

Tissue distribution

Analysis of northern blots from a variety of human and murine tissues (Clontech) revealed *IL21R* transcripts only in lymphoid tissues, including spleen, thymus, peripheral blood lymphocytes (PBLs) and lymph node (Fig. 2). *In situ* hybridization using antisense probes specific to two different regions of *IL21R* resulted in positive signals in a subset of cells within normal tonsil, lymph node and thymus, and in the lamina propria of duodenal villi, as well as in fibrotic (but not normal) lung (data not shown).

To characterize the specific cell types expressing surface IL21R, we carried out flow cytometry on resting and activated peripheral B, T

and NK cells using biotinylated IL21. Resting CD23⁺ B cells from three different donors show positive staining, which is blocked in each case by 15-fold excess unlabelled IL21 (Fig. 3a). Although receptor expression was not consistently detectable by flow cytometry on peripheral NK cells, we were able to detect it on the NK cell line NK-92 (ref. 14) (Fig. 3b), and on fresh NK cells by polymerase chain reaction with reverse transcription (RT-PCR) (data not shown). We carried out flow cytometry on a series of mouse and human cell lines. In addition to NK-92, cell lines found to express IL21R included Raji (human Burkitt lymphoma), Jurkat (human T-cell leukaemia), IM-9 (human B-cell line), HS Sultan (human B-cell line) and A20 (mouse B-cell lymphoma). IL21R expression was not detected on EL4 (mouse thymoma) or K562 (human T-cell leukaemia) (Fig. 3b; and data not shown).

Expression of *IL21* ligand in normal tissues was not detectable by northern analysis (data not shown); however, the sources from which human and mouse *IL21* cDNAs were cloned suggested that it is probably produced by activated peripheral T cells. Quantitative RT-PCR analysis was used to measure activation-induced changes in *IL21* message levels in purified human peripheral T cells. As CD3⁺ T cells express *IL21* transcripts on stimulation with PMA and ionomycin, CD4⁺ and CD8⁺ T-cell subsets were isolated and stimulated either with PMA plus ionomycin or with an immobilized anti-CD3 antibody. CD4⁺ T cells are stimulated to express *IL21* under both sets of conditions; however, expression is significantly

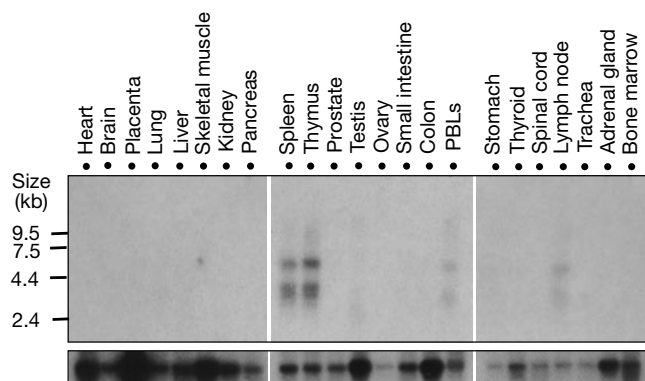


Figure 2 Expression analysis of IL21R in normal tissues. Northern analysis was performed on multiple-tissue northern blots (human, human III, human IV; Clontech). Filters were exposed for 48 h at -80°C . Bottom panel shows hybridization with a human transferrin receptor probe.

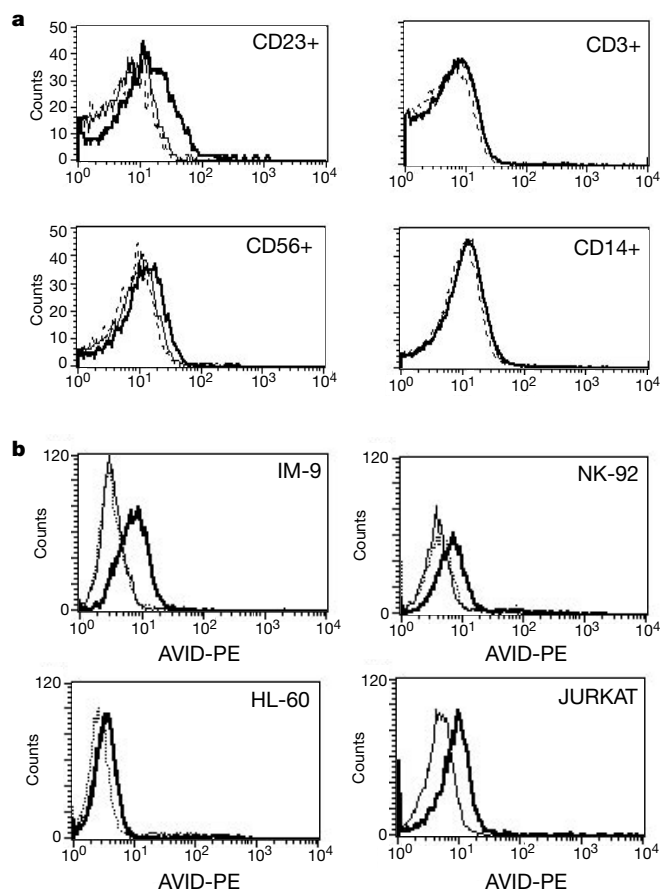


Figure 3 Flow cytometric analysis of IL21R expression. Human peripheral blood lymphocytes (a) or cell lines (b) were stained with antibodies to the markers indicated, and/or with $2\text{ }\mu\text{g ml}^{-1}$ biotinylated IL21 (bold lines). Binding was competed using $30\text{ }\mu\text{g ml}^{-1}$ unlabelled IL21 (dashed lines). Thin lines represent secondary reagent only (Streptavidin APC). Other reagents used are anti-CD23-PE, anti-CD3-CyChrome, anti-CD14-PE and anti-CD56-FITC. Results shown in a are representative of three different donors.

higher in the samples stimulated with anti-CD3 (Fig. 4a). To further define the conditions that induce *IL21* expression, we purified peripheral T cells from a different donor and stimulated with anti-CD3 either alone or in combination with anti-CD28. The combination of anti-CD3 and anti-CD28—conditions that best mimic an antigen encounter *in vivo*—provided higher induction of *IL21* expression than that seen with anti-CD3 alone in this assay (Fig. 4b). The small increase in message seen in stimulated CD8⁺ cells in both experiments is probably due to contaminating CD4⁺ cells (1–3%). CD19⁺ B cells and CD14⁺ monocytes did not express *IL21* under any stimulation condition (data not shown).

Biological activities

Because resting B lymphocytes express IL21R (see Fig. 3a), we designed assays to determine the effects of IL21 on B-cell proliferation. Although IL21 alone has no mitogenic effect on isolated human B cells (data not shown), it significantly co-stimulates proliferation induced by anti-CD40 monoclonal antibodies (Fig. 5a). In contrast, IL21 inhibits proliferation induced by the combination of plate-bound anti-IgM antibodies and IL4 (Fig. 5b). Both effects are reversed by the addition of sIL21R to the culture.

We also tested IL21 in T-cell proliferation assays and found that it co-stimulates anti-CD3-activated murine thymocytes (Fig. 6a), leading to an accelerated outgrowth of CD8⁺CD4⁺ cells (data not shown). We did not observe significant levels of proliferation of thymocytes in response to IL21 in the absence of anti-CD3. IL21 also co-stimulates anti-CD3-activated mature murine (Fig. 6b) and human (Fig. 6c) T cells. In each case, addition of sIL21R blocked this IL21-dependent proliferation (data not shown). Notably, our early experiments suggested that the proliferative effect of IL21 is subtler on human T cells than on murine T cells. We reasoned that this might be a function of the prior activation state of the human T cells, as murine T cells are less likely to be pre-activated *in vivo*. Indeed, when we separated CD45RA⁺ (naïve) and CD45RO⁺ (memory) human T cells and assayed them for proliferative responses to IL21, we found that naïve, but not memory, T cells were co-stimulated by IL21 (Fig. 6c; and data not shown). To compare the activity of IL21 to that of related cytokines on murine T cells, we performed a cross-titration of IL21 and IL2, IL7 or IL15, in the presence or absence of a suboptimal concentration of immobilized anti-CD3 monoclonal antibody (Fig. 6d). Unlike its related counterparts, IL21 has no significant proliferative effect on T cells in the absence of anti-CD3 or other stimuli (Fig. 6d; and data not shown). However, it does act in concert with IL2, IL15, and, to a lesser extent, with IL7 to enhance T-cell proliferation, either with or without anti-CD3 stimulation (Fig. 6d).

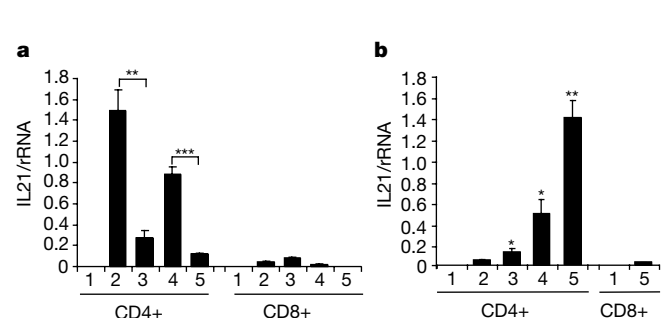


Figure 4 Quantitative RT-PCR analysis of *IL21* expression in lymphocyte subsets. **a**, Stimulation conditions are resting (lane 1); 3.5 h anti-CD3 (lane 2) or PMA plus ionomycin (lane 3); 16 h anti-CD3 (lane 4) or PMA plus ionomycin (lane 5). **b**, Stimulation conditions are resting (lane 1); 2 h anti-CD3 (lane 2); and anti-CD3 plus anti-CD28 for 2 h (lane 3), 4 h (lane 4) or 16 h (lane 5). Vertical bars represent the ratio of *IL21* message in each sample to that of 18S ribosomal RNA in the same sample. Values plotted represent the mean ($\pm$ s.d.) obtained from triplicate assays. Asterisk, $P < 0.05$; two asterisks, $P < 0.005$; three asterisks, $P < 0.001$.

To assess the activity of IL21 on lymphoid progenitors, we co-cultured human non-adherent bone marrow progenitor cells with stem cell factor and human IL21 in combination with various other cytokines. Although IL21 alone does not induce significant proliferation of marrow progenitors, it does synergize with IL15, consistently causing at least a twofold increase in cell proliferation as compared with cultures containing stem cell factor and IL15 alone (data not shown). Flow cytometric analysis of these expanded cells for lineage-specific markers revealed expression of the NK cell marker CD56 (data not shown).

Other groups have shown that CD3⁺CD56⁺ NK cells can be generated from 21-day cultures of human CD34⁺ haematopoietic progenitor cells (HPCs) supplemented with IL15 in the absence of stromal cells, an effect that is enhanced by the addition of Flt3L^{15,16}. Given our observation that CD56⁺ NK cells grow out of bone marrow cultures stimulated with IL15 and IL21, we undertook an analysis of the effect of IL21 on the development of NK cells from HPCs. We isolated CD34⁺ human HPCs (95% pure) and cultured them either with Flt3L with and without IL21, or with Flt3L plus IL15 with and without IL21. Cells were collected on day 15 to assess their NK markers by flow cytometry (Fig. 7a) and also their capacity to lyse MHC⁻ K562 cells in a standard ⁵¹Cr-release assay (Fig. 7b). As previously reported¹⁶, the combination of IL15 and Flt3L alone induced the outgrowth of CD56⁺ cells. In contrast, IL21 and Flt3L alone did not significantly expand the CD56⁺ fraction. In cultures exposed to the combination of Flt3L, IL15 and IL21, however, there was a significant increase in total cell numbers and the proportion of cells that were CD56⁺CD16^{bright}, when compared to cultures supplemented with Flt3L and IL15 alone. The cells growing out of these Flt3L, IL15, IL21 cultures were differentiated NK cells, as they were large and granular and expressed both CD56 and CD16, but did not express CD3 (Fig. 7a; and data not shown). Furthermore, these cells exhibited significantly higher effector function at day 15 than those cells grown with only IL15 and Flt3L (Fig. 7b). After 21 days in culture, however, the cells cultured in the presence of Flt3L and IL15 (without IL21) had gained lytic activity, but remained CD16^{low/-}, consistent with previous reports^{15,16} (data not shown).

We also tested whether IL21 can act on mature NK cells. We isolated CD56⁺ cells from fresh human peripheral blood mononuclear cell (PBMCs) (Fig. 7c), and cultured them for 1–2 days with and without 30 ng ml⁻¹ IL2, IL15 or IL21. The cells cultured with IL21 exhibited enhanced lytic activity on K562 target cells, although the effect of IL21 was never quite as pronounced as that of IL2 or IL15 (Fig. 7d). Treatment of NK cells with combinations of IL21 and

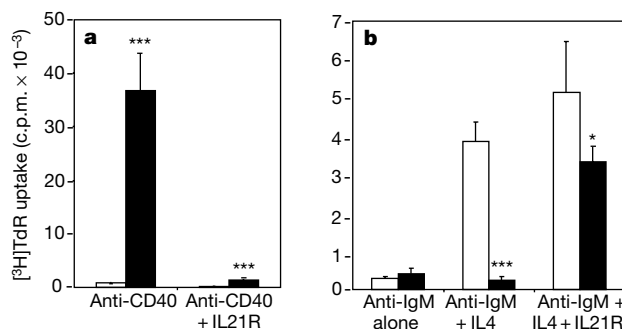


Figure 5 Effect of IL21 on the proliferative responses of human B cells. B cells were purified from peripheral blood of normal human donors and cultured either in the absence (open bars) or in the presence (filled bars) of 10 ng ml⁻¹ IL21. **a**, Anti-CD40 and 25 μ g ml⁻¹ sIL21R were used as additional stimuli as indicated. **b**, Combinations of immobilized anti-IgM, IL4 and 25 μ g ml⁻¹ sIL21R were used as additional stimuli as indicated. Values plotted represent the mean value ($\pm$ s.d.) obtained from triplicate cultures. Similar results were obtained in four independent experiments. Asterisk, $P < 0.05$; two asterisks, $P < 0.005$; three asterisks, $P < 0.001$.

IL2 or IL15 led to an additive effect on lytic function (data not shown).

Discussion

We have cloned a cytokine–receptor pair, interleukin-21 and IL21R, that both regulates the proliferation of mature B and T cells in response to activating stimuli, and mediates expansion of NK cell populations from bone marrow. *IL21R* is expressed in peripheral B cells as well as in cell lines of B, T and NK lineages. *IL21* is expressed in a highly controlled manner, by CD4⁺ T cells that have been activated using stimuli that closely mimic an antigen encounter.

Given the structural similarity between IL21 and the IL15/IL4/IL2 family of cytokines, together with the structural relationship between IL21R and both IL2R β and IL4R α , it is reasonable to speculate that, although the IL21R receptor appears to be capable of signal transduction through homodimerization (as is IL4R α ¹⁷), it may physiologically associate in a heteromeric receptor with IL2R γ c. The IL2R γ c receptor subunit participates in forming the receptors for IL15, IL4, IL7, IL9 and IL2 (for review, see ref. 18). Our preliminary data suggest that IL2R γ c may indeed be a part of the IL21 receptor complex; further characterization is in progress.

IL21 promotes expansion of mature B cells in response to stimulation through CD40. Interactions between CD40 and its

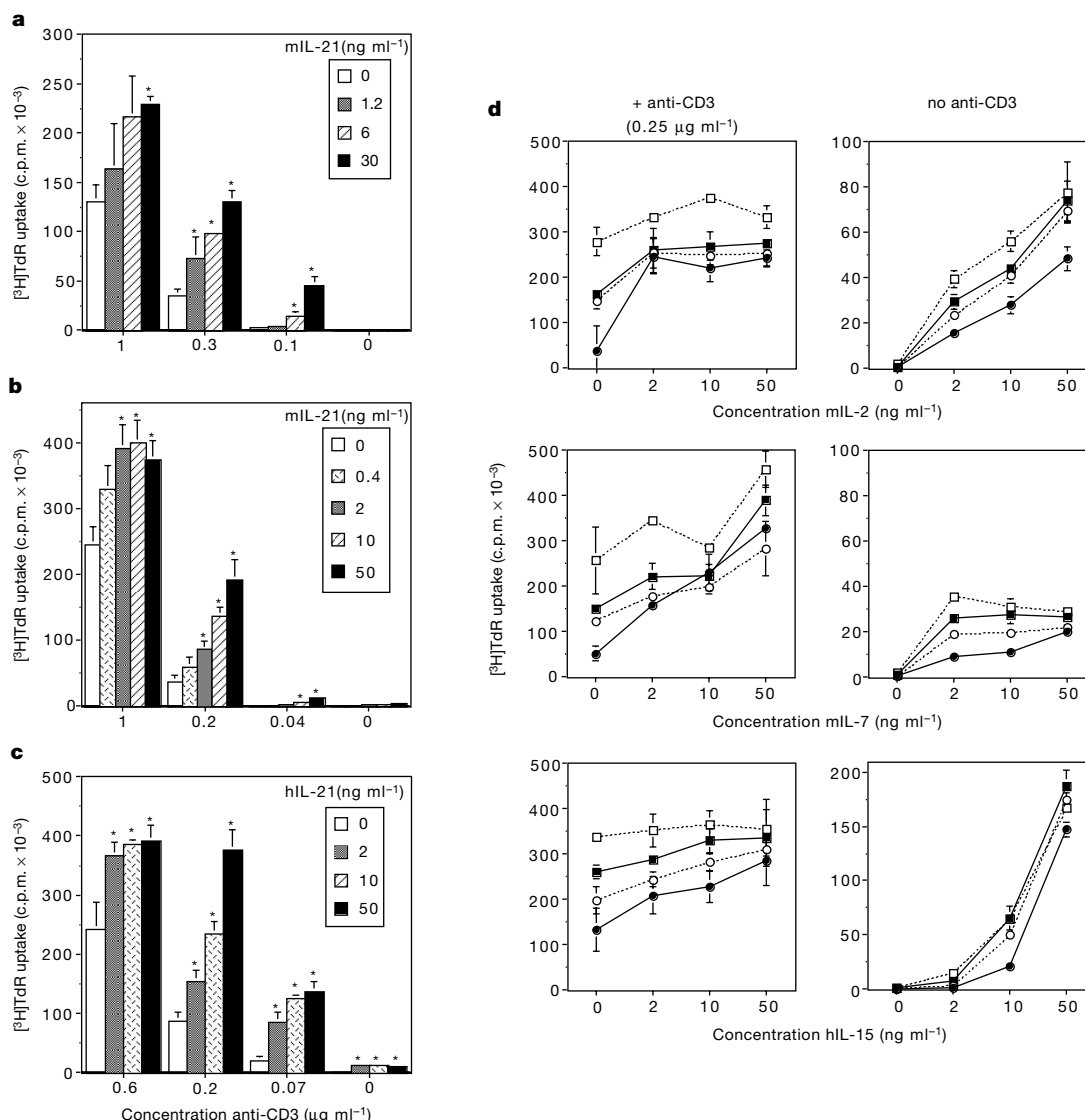


Figure 6 IL21 co-stimulates anti-CD3 activated thymocytes and mature T cells. **a**, Whole C57Bl/6 thymocytes were assayed for proliferation in response to plate-bound anti-CD3 ϵ monoclonal antibody in the absence or presence of increasing concentrations of mL21, as indicated. Values plotted represent the mean value ($\pm$ s.d.) obtained from triplicate cultures. Asterisks indicate significant values ($P < 0.05$ by Student's *t*-test) above those from wells lacking IL21. Data shown are representative of results from six experiments. Thymocytes stimulated with 2 ng ml^{-1} PMA and 500 ng ml^{-1} ionomycin as a positive control incorporated 445×10^3 c.p.m. in the same time period. **b**, Mature murine T cells were enriched from pooled C57Bl/6 spleen and lymph nodes by depletion of CD19⁺ B cells. The resulting cell population (78% CD3⁺ and 3% B220⁺) was assayed as in **a**. T cells stimulated with PMA and ionomycin incorporated 625×10^3 c.p.m. Results shown are representative of five experiments. **c**, Human CD45RA⁺ T cells were isolated as

described in Methods. These enriched naive T cells (75% CD3⁺CD45RA⁺, 1% CD3⁺CD45RO⁺, 4% CD11b⁺, 5% CD20⁺) were tested in triplicate cultures for proliferation in a 3-day assay to immobilized anti-human CD3 monoclonal antibody, in the presence or absence of human IL21, as indicated. Control T cells stimulated with PMA and ionomycin incorporated an average of 615×10^3 c.p.m. Results shown are representative of three experiments. **d**, Mature murine T cells were purified as in **b** and incubated in the absence (right) or presence (left) of 0.25 $\mu\text{g ml}^{-1}$ anti-CD3 ϵ monoclonal antibodies, along with increasing amounts of mL2 (top), mL7 (middle) or hIL15 (bottom) in a 3-day proliferation assay. A cross-titration of 0 (filled circles), 2 (open circles), 10 (filled squares) or 50 ng ml^{-1} mL21 was also included, to test for synergy and/or competition among the related cytokines. Results shown are representative of three experiments.

ligand, CD40L, are an integral part of the communication between B cells and helper T cells in mounting thymus-dependent (TD) immune responses, and are vital in generation of memory B cells and germinal centres (for reviews, see refs 19, 20). Defects in CD40L lead to X-linked hyper-IgM syndrome, a primary immunodeficiency with characteristics including absence of germinal centres and increased incidence of autoimmunity (for review, see ref. 21). In contrast to its synergistic effect with CD40 stimulation, IL21 completely inhibits proliferation of B cells stimulated with anti-IgM and IL4. This suggests that signalling through IL21R can differ markedly, depending on the co-stimuli encountered by the B cell during an immune response.

In addition to its effects on B cells, IL21 promotes enhanced expansion of T cells co-stimulated with anti-CD3. In humans, the response to IL21 appears to be limited to naive CD45RA⁺ T cells, suggesting either that IL21R is downregulated on activated/memory T cells, or that the signalling cascade downstream of IL21R in naive compared with memory T cells is distinct. As activated CD4⁺ T cells also produce IL21, this cytokine may have an autocrine/paracrine

role in regulation of the immune response by promoting the expansion of helper T cells and by eliciting the appropriate responses from neighbouring activated B cells.

IL21 promotes expansion and differentiation of NK cells from bone marrow progenitors *in vitro*, in synergy with Flt3L and IL15. The most striking result is the development of CD16^{bright}, highly lytic NK cells by 15 days in culture in the presence of IL21. In agreement with previous reports^{15,16}, we also observed the development of lytic CD16^{low/-} cells from CD34⁺ marrow cells after 21 days in culture in the presence of only Flt3L and IL15 (data not shown). CD16 is thought to mediate antibody-dependent cellular cytotoxicity (for review, see ref. 22), and its expression on IL21-treated NK cells reflects an enhanced maturational state.

NK cells are lymphoid cells that participate in the innate immune system's surveillance of malignant cells and thus may help to control metastatic cancer. These cells are also essential for natural immunity against microbial pathogens, in particular viruses; they influence the graft-versus-leukaemia effect after bone marrow transplantation; and they are involved in the regulation of haematopoiesis (for

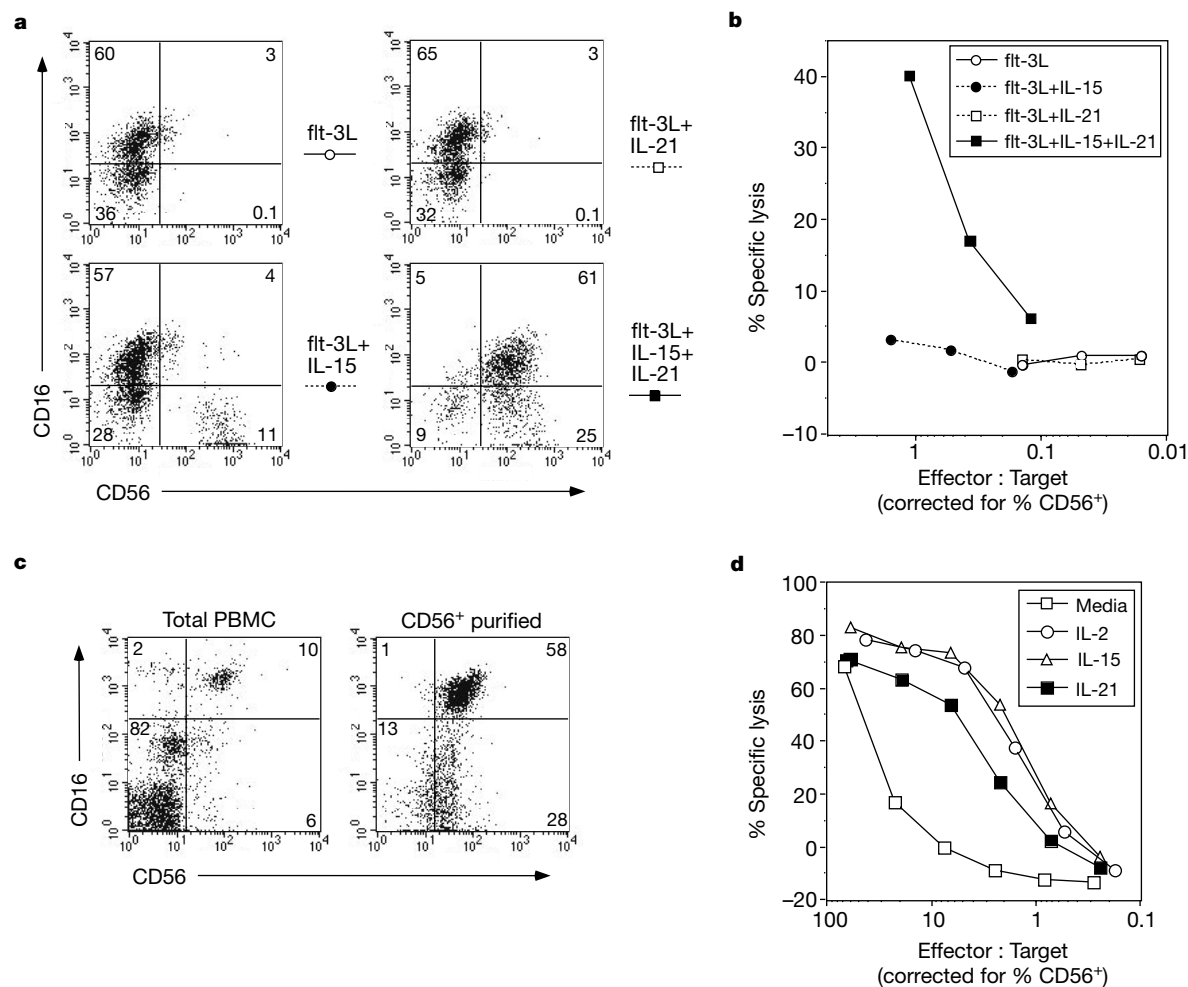


Figure 7 IL21 synergizes with IL15 and Flt3L to enhance the proliferation and differentiation of NK cells from CD34⁺ bone marrow progenitors, and alone enhances the effector function of mature NK cells. **a**, Human CD34⁺ bone marrow cells (95% pure) were cultured for 15 days in the presence of 2 ng ml⁻¹ Flt3L (open circles), Flt3L plus 20 ng ml⁻¹ hIL15 (filled circles), Flt3L plus 15 ng ml⁻¹ hIL21 (open squares) or Flt3L plus hIL15 plus hIL21 (filled squares). An aliquot of cells from each culture condition was stained and analysed by flow cytometry for surface expression of CD56 and CD16 (all cells were CD3⁻). Data were gated on viable cells based on their light scatter characteristics; the percentages of cells in each quadrant are indicated. Results shown are representative of four experiments using four different bone marrow donors. **b**, Cells shown in **a** were collected, washed and assayed for their ability to lyse ⁵¹Cr-labelled K562 (MHC⁻) target

cells in a standard 4-h assay. The effector-to-target ratio was corrected for the percentage of CD56⁺ cells in each culture. **c**, Mature CD56⁺ NK cells were purified from fresh human PBMCs by positive selection, stained with anti-CD56-PE and anti-CD16-CyChrome monoclonal antibodies, and analysed for purity by flow cytometry. **d**, Purified CD56⁺ NK cells shown in **c** were cultured in media alone (open squares), or with 30 ng ml⁻¹ of hIL2 (circles), hIL15 (triangles) or hIL21 (filled squares) for 24 h, and then tested for lytic function in the assay described in **b**. No significant differences were noted in the number of cells recovered from each culture condition. Similar results were achieved when the cells were cultured for 48 h before assay. Results shown are representative of five experiments using four different blood donors.

reviews, see refs 22–25). The contrasting effects of IL21 on the proliferation of B cells, depending on their mode of activation, as well as its effects on NK cell maturation and on T cells, suggest that it may be an important regulator of the immune system.

Note added in proof: Following acceptance of this article, a paper was published describing the receptor²⁹. □

Methods

Library construction

Human CD3⁺ PBMNC and Balb/C CD90⁺ splenocytes were activated using 10 ng ml⁻¹ PMA and 0.5 µg ml⁻¹ ionomycin (Calbiochem). We isolated RNA using Rneasy Midi Kit (Qiagen) and mRNA using MPG mRNA purification kit (CPG Inc.). cDNA was synthesized using a modified Gubler–Hoffman method²⁶ and cloned into expression vector pZPNX.

Protein expression and purification

The extracellular portion of IL21R was cloned into an expression vector with a C-terminal Glu–Glu (EE) tag and expressed in BHK cells. Protein was purified from conditioned media using anti-EE G-sepharose and POROS HQ-50 columns (PerSeptive BioSystems), and then Sephadex S200 (Pharmacia) and ActiClean Etox (Sterogene) columns. We cloned human IL21 into pDC312 (ref. 27) and expressed it in suspension-adapted CHO DG44 cells. Mouse IL21 was cloned into pZP9 and expressed in BHK cells. Conditioned media were separated using Poros 50 HS columns (PerSeptive BioSystems). Human IL21 was then purified on a Sephacryl S-200 column (Pharmacia). Mouse IL21 was purified using Poros AL/IL21R columns, and then Sephacryl S-200.

Real-time PCR

Human PBMCs were isolated from fresh whole blood. We purified CD4⁺ and CD8⁺ cells by positive selection using a MACS magnetic separation column (Miltenyi Biotec). Purity was assessed by flow cytometry; CD4⁺ cells were 85% and 97% pure, and CD8⁺ cells were 84% and 81% pure in the assays shown in Fig. 4a and b, respectively. We carried out real-time quantitative RT–PCR using the ABI PRISM 7700 sequence detection system (PE Applied Biosystems). Primers were (5′-tgtagatgac ttgtccctg aa-3′) and (5′-aacaggaacaa agctgaccac tca-3′). The TaqMan probe (5′-tctgcagctt ccagaagatg tagagacaaa c-3′) was synthesized by PE Applied Biosystems. Levels of IL21 mRNA were determined by analysis of total RNA samples using the one-step RT–PCR method (PE Applied Biosystems). RNA from BHK cells expressing human IL21 was used to generate a standard curve for each assay. Relative IL21 mRNA levels were determined by the standard curve method (user bulletin no. 2, PE Applied Biosystems) using 18S ribosomal RNA (TaqMan ribosomal RNA control reagents protocol, PE Applied Biosystems) measurements to normalize.

Purification of bone marrow HPCs

We prepared CD34⁺ cells from fresh human bone marrow mononuclear cell preparations (Poietics Technologies) using a CD34 magnetic selection system (Dyna). Mononuclear cells were washed, resuspended in PBS/2% BSA, and incubated with anti-CD34 magnetic beads for 30 min at 4 °C. After washing to remove non-rosetted cells, the CD34⁺ cells were released by incubation with polyclonal antibody to the Fab portion of the CD34 monoclonal, 15 min at 37 °C. Cells purified by this technique have an average purity of 95%, as assessed by flow cytometry.

NK cytotoxicity assays

Aliquots of bone marrow cultures were stained for flow cytometric analysis with anti-CD3-FITC, anti-CD56-PE and anti-CD16-CyChrome monoclonal antibodies (PharMingen), and analysed on a FACSCalibur using CellQuest software (Becton Dickinson). We examined NK-mediated target cytotoxicity by standard ⁵¹Cr-release assay. Effector cells were prepared from human CD34⁺ bone marrow cells cultured for 15–21 days, or human CD56⁺ PBMCs cultured for 1–2 days, with various combinations of cytokines, and then collected, washed, counted, mixed at various concentrations with 5 × 10³ ⁵¹Cr-labelled K562 cells in 96-well round-bottomed plates, and incubated for 4 h at 37 °C. After incubation, half the supernatant from each well was collected for determination of radioactivity. The percentage of specific ⁵¹Cr release was calculated from the formula 100 × (X–Y)/(Z–Y), where X is ⁵¹Cr release in the presence of effector cells, Y is the spontaneous release in the absence of effectors, and Z is the total ⁵¹Cr release from target cells incubated with 0.5% Triton X-100.

T-cell proliferation assays

For mouse assays, T cells from C57BL/6 mice were isolated from pooled splenocytes and lymph node lymphocytes and depleted of CD19⁺ B cells using a MACS column (Miltenyi Biotec). Whole thymocytes were also collected. Cells were cultured in triplicate with murine IL21 (0–30 ng ml⁻¹) in 96-well flat-bottomed plates (pre-coated with anti-CD3 monoclonal antibody 2C11; PharMingen) for 3 d at 37 °C. Each well was pulsed with 1 µCi [³H]thymidine on day 2, and plates were collected and counted 16 h later. For human assays, whole blood was collected from a healthy donor and depleted of red blood cells using Ficoll Paque gradients (Amersham Pharmacia). To enrich for naive T cells, CD11b⁺ monocytes and CD19⁺ B cells were first depleted using a MACS column, and then positively selected for CD45RA⁺ cells. These T cells were cultured in triplicate with human IL21 (0–50 ng ml⁻¹) in 96-well flat-bottomed plates (pre-coated where indicated with anti-human CD3 monoclonal antibody UCHT1; PharMingen) for 3–5 days at 37 °C.

Each well was pulsed with 1 µCi [³H]thymidine on day 2–4, and plates were collected and counted 16–18 h later.

B-cell proliferation assays

We purified human peripheral B cells with anti-CD19 magnetic beads (Miltenyi Biotec). For anti-IgM stimulation, plates were pre-coated with anti-human IgM (Southern Biotech Associates) and cultures were supplemented with 10 ng ml⁻¹ IL4 (R&D Systems). Anti-human CD40 monoclonal antibody (Genzyme) was added to cultures at 1 µg ml⁻¹. Proliferation was assessed by the incorporation of [³H]thymidine (1 µCi per well) during the last 16 h of a 120-h culture. We carried out treatments in triplicate.

Received 19 July; accepted 13 September 2000.

- Wlodawer, A., Pavlovsky, A. & Gustchina, A. Hematopoietic cytokines: similarities and differences in the structures, with implications for receptor binding. *Protein Sci.* **2**, 1373–1382 (1993).
- Cosman, D. The hematopoietin receptor superfamily. *Cytokine* **5**, 95–106 (1993).
- Lok, S. *et al.* Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production *in vivo*. *Nature* **369**, 565–568 (1994).
- Palacios, R. & Steinmetz, M. IL3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes *in vivo*. *Cell* **41**, 727–734 (1985).
- Bazan, J. F. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl Acad. Sci. USA* **87**, 6934–6938 (1990).
- Vigon, I. *et al.* Characterization of the murine *Mpl* proto-oncogene, a member of the hematopoietic cytokine receptor family: molecular cloning, chromosomal location and evidence for a function in cell growth. *Oncogene* **8**, 2607–2615 (1993).
- Skoda, R. C. *et al.* Murine c-mpl: a member of the hematopoietic growth factor receptor superfamily that transduces a proliferative signal. *EMBO J.* **12**, 2645–2653 (1993).
- Murakami, M. *et al.* Critical cytoplasmic region of the interleukin 6 signal transducer gp130 is conserved in the cytokine receptor family. *Proc. Natl Acad. Sci. USA* **88**, 11349–11353 (1991).
- Drachman, J. G. & Kaushansky, K. Dissecting the thrombopoietin receptor: Functional elements of the *Mpl* cytoplasmic domain. *Proc. Natl Acad. Sci. USA* **94**, 2350–2355 (1997).
- Gurney, A. L., Wong, S. C., Henzel, W. J. & De Sauvage, F. J. Distinct regions of c-mpl cytoplasmic domain are coupled to the JAK-STAT signal transduction pathway and Shc phosphorylation. *Proc. Natl Acad. Sci. USA* **92**, 5292–5296 (1995).
- von Heijne, G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
- Hage, T., Sebald, W. & Reinemer, P. Crystal structure of the interleukin-4/receptor α chain complex reveals a mosaic binding interface. *Cell* **97**, 271–281 (1999).
- Zurawski, S. M. *et al.* Definition and spatial location of mouse interleukin-2 residues that interact with its heterotrimeric receptor. *EMBO J.* **12**, 5113–5119 (1993).
- Gong, J. H., Maki, G. & Klingemann, H. G. Characterization of a human cell line (NK-92) with phenotypic and functional characteristics of activated natural killer cells. *Leukemia* **8**, 652–658 (1994).
- Mrozek, E., Anderson, P. & Caligiuri, M. A. Role of interleukin-15 in the development of human CD56⁺ natural killer cells from CD34⁺ hematopoietic progenitor cells. *Blood* **87**, 2632–2640 (1996).
- Yu, H. *et al.* Flt3 ligand promotes the generation of a distinct CD34⁺ human natural killer cell progenitor that responds to interleukin-15. *Blood* **92**, 3647–3657 (1998).
- Kammer, W. *et al.* Homodimerization of interleukin-4 receptor α chain can induce intracellular signaling. *J. Biol. Chem.* **271**, 23634–23637 (1996).
- Sugamura, K. *et al.* The interleukin-2 receptor γ chain: Its role in the multiple cytokine receptor complexes and T cell development in XSCID. *Annu. Rev. Immunol.* **14**, 179–205 (1996).
- Grewal, I. S. & Flavell, R. A. CD40 and CD154 in cell-mediated immunity. *Annu. Rev. Immunol.* **16**, 111–135 (1998).
- Foy, T. M., Aruffo, A., Bajorath, J., Buhlmann, J. E. & Noelle, R. J. Immune regulation by CD40 and its ligand gp39. *Annu. Rev. Immunol.* **14**, 591–617 (1996).
- Callard, R. E., Armitage, R. J., Fanslow, W. C. & Spriggs, M. K. CD40 ligand and its role in X-linked hyper-IgM syndrome. *Immunol. Today* **14**, 559–564 (1993).
- Robertson, M. J. & Ritz, J. Biology and clinical relevance of human natural killer cells. *Blood* **76**, 2421–2438 (1990).
- Whiteside, T. L., Vujanovic, N. L. & Herberman, R. B. Natural killer cells and tumor therapy. *Curr. Top. Microbiol. Immunol.* **230**, 221–244 (1998).
- Biron, C. A., Nguyen, K. B., Pien, G. C., Cousens, L. P. & Salazar-Mather, T. P. Natural killer cells in antiviral defense: function and regulation by innate cytokines. *Annu. Rev. Immunol.* **17**, 189–220 (1999).
- Williams, N. S. *et al.* Natural killer cell differentiation: insights from knockout and transgenic mouse models and *in vitro* systems. *Immunol. Rev.* **165**, 47–61 (1998).
- Gubler, U. & Hoffman, B. J. A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269 (1983).
- Morris, A. *et al.* in *Animal Cell Technology* (ed. Carrondo, M. J. T.) 529–534 (Kluwer Academic, The Netherlands, 1997).
- Stahl, N. *et al.* Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349–1353 (1995).
- Dzaki, K., Kikly, K., Michalovich, D., Young, P. R. & Leonard, W. J. Cloning of a type I cytokine receptor most related to the IL-2 receptor β chain. *Proc. Natl Acad. Sci. USA* **97**, 11439–11444 (2000).

Acknowledgements

We thank J. Lehner, J. Rosser, S. McMillen, S. Matthews, J. Coleman, S. Sexson, A. Adan, W. Lint, K. Hodges, L. Smith, J. Rodriguez, C. Noriega, B. Dedinsky, K. Kontor, J. Gray, M. Rixon and R. Hoffman for technical assistance; and P. McKernan, J. Kelly, S. Jaspers, G. McKnight, G. Rosenberg, K. Swiderek, D. Prunkard, B. Persson, W. Kindsvogel and T. Bukowski for project advice. Nomenclature has been approved by HUGO.

Correspondence and requests for materials should be addressed to D.F. (e-mail: fosterd@zgi.com). Sequences described have been submitted to GenBank under accession nos AF254067, AF254068, AF254069 and AF254070.

A photorefractive organically modified silica glass with high optical gain

Pavel Cheben^{*}, Francisco del Monte^{†‡}, Dennis J. Worsfold[§], Dave J. Carlsson[§], Chander P. Grover^{*} & John D. Mackenzie[†]

^{*} Institute for National Measurement Standards, National Research Council, K1A 0R6 Ottawa, Canada

[†] Department of Materials Science and Engineering, University of California, Los Angeles, California 90095, USA

[‡] Instituto de Ciencia de Materiales de Madrid, Consejo Superior de Investigaciones Científicas, Cantoblanco, 28049 Madrid, Spain

[§] Institute for Chemical Process and Environmental Technology, National Research Council, K1A 0R6 Ottawa, Canada

Photorefractive materials¹ exhibit a spatial modulation of the refractive index due to redistribution of photogenerated charges in an optically nonlinear medium. As such, they have the ability to manipulate light and are potentially important for optical applications¹ including image processing, optical storage, programmable optical interconnects and simulation of neural networks. Photorefractive materials are generally crystals, polymers and glasses with electro-optic or birefringent properties and non-centrosymmetric structure². Here we report the photorefractive effect in both non-centrosymmetric and centrosymmetric azo-dye-doped silica glasses, in which refractive index gratings that are spatially phase-shifted with respect to the incident light intensity pattern are observed. The effect results from a non-local response of the material to optical illumination, and enables the transfer of energy between two interfering light beams (asymmetric two-beam coupling). Although the writing time for the present grating is relatively slow, we have achieved a two-beam coupling optical gain of 188 cm⁻¹ in the centrosymmetric glasses, and a gain of 444 cm⁻¹ in the non-centrosymmetric structures. The latter are fabricated using a corona discharge process³ to induce a permanent arrangement of azo-dye chromophores.

Our investigations were performed on an azo-dye-doped silica glass prepared by the sol-gel process. The sol-gel technique has been widely used in the preparation of transparent oxide glasses by hydrolysis and condensation of metal alkoxides⁴ since the first synthesis of silica from silicon alkoxide was reported by Ebelmen in 1844 (ref. 5). Given that in the sol-gel process little or no heating is required, organic molecules with low thermal stability can be incorporated into inorganic ceramic or glassy hosts. In our experiment, the silica glass is doped with organic molecules in order to provide the properties necessary for photorefractivity such as photosensitivity, charge transport and nonlinear optical properties (Fig. 1). A small amount of 2,4,7-trinitro-9-fluorenone (TNF) is used to increase the photosensitivity in the visible region. *N*-ethylcarbazole (ECZ), which forms a charge-transfer complex with TNF, is used as the charge-transporting part. Both the TNF and ECZ are present as 'guests' in the glass without being covalently attached to the silica glass network.

The azo-dye 2,5-dimethyl-4-(2-hydroxyethoxy)-4'-nitroazobenzene (DMHNAB) is used as a nonlinear optical chromophore and is structurally related to 2,5-dimethyl-4-(*p*-nitrophenylazo)anisole (DMNPAA), which has previously been used in a high-optical-gain photorefractive polymer⁶. The chromophore molecules are covalently bound to the silica glass backbone in order to achieve the high dye concentration required for efficient nonlinear optical properties, while avoiding the dye crystallization that is typically observed in guest-host photorefractive polymers with a high content of non-bonded dye. DMHNAB was attached to a sol-gel

precursor, 3-(isocyanatopropyl)triethoxysilane (ICPTES), using an analogous technique to that reported for ICPTES functionalization with azobenzene dye disperse red 1 (ref. 7). The completion of the reaction between the dye and ICPTES was confirmed by infrared spectra and nuclear magnetic resonance measurements. We obtained DMHNAB-gel by copolymerization of DMHNAB-ICPTES and tetraethoxysilane (TEOS) precursors. The sol-gel precursors were mixed in pyridine solvent with deionized water, ECZ, TNF, and hydrochloric acid (HCl) as a polymerization catalyst. After stirring the solution of pyridine(11):DMHNAB-ICPTES(1):ECZ(1.6):TEOS(1.1):H₂O(11):HCl(0.6):TNF(0.002) (bracketed numbers indicate molar equivalents) for 30 minutes at 65 °C, 0.1 ml of the solution was cast onto a glass plate coated with a transparent indium tin oxide electrode. After heating at 60 °C for 4 hours, we obtained a soft, transparent gel film of approximately 30 µm thickness.

There is a strong need for a permanently poled photorefractive organic material with a high glass transition temperature, for long-term holographic data storage⁸. We prepared a permanently poled glass by heating the soft gel to 125 °C for 30 minutes, in a corona discharge that was generated between a sharp needle electrode at 5 kV static potential and the sample held 1 cm below. In the external electric field produced by charges deposited on the film surface through the corona process³, chromophores tend to orient their dipole moments parallel to the field gradient; non-centrosymmetry, which is required for electro-optic properties, is induced. At the same time the orientational mobility of the chromophore is decreased by heat-induced densification of the gel. The orientation persists and remains stable after the poling field is removed and the temperature is decreased to ambient temperature. After completion of corona poling and gel densification we obtained a hard glassy film. The densification process is essential for slowing down diffusive randomization of chromophore alignment and for improving the mechanical, electrical and thermal properties of the film, given that the poled glass is an intrinsically unstable, nonequilibrium thermodynamic system. We also prepared a centrosymmetric glass film by heat-induced gel densification in the absence of the external electric field. The final composition of the cured samples was determined by Fourier-transform infrared (FTIR) spectroscopy, which showed the presence of some SiOH groups, and gave the following composition: DMHNAB-urethane-SiO_{1.5}(1):SiO_{1.5}OH(1.1):ECZ(0.2):TNF(0.002) (in molar equivalents).

To confirm the photorefractive properties of our glass, it is essential to distinguish between the photorefractive effect and other mechanisms of grating formation such as photochromism, electrochromism, thermorefractive, generation of excited states and conventional electronic third-order susceptibility⁹. To do this, we performed a two-beam-coupling (2BC) experiment. The transmitted intensity of the writing beam 1 was monitored while the

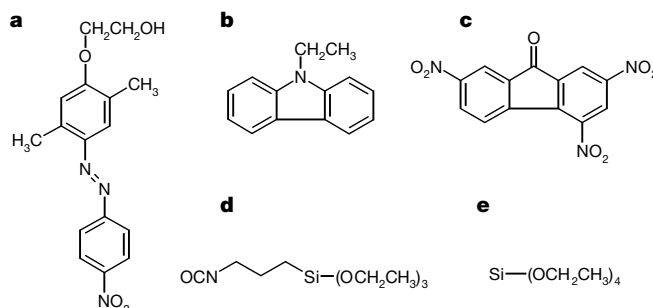


Figure 1 Chemical structure of the organic dopants and sol-gel precursors used in the synthesis of the photorefractive glass. **a**, 2,5-dimethyl-4-(2-hydroxyethoxy)-4'-nitroazobenzene (DMHNAB); **b**, *N*-ethylcarbazole (ECZ); **c**, 2,4,7-trinitro-9-fluorenone (TNF); **d**, 3-(isocyanatopropyl)triethoxysilane (ICPTES); **e**, tetraethoxysilane (TEOS).

mutually coherent writing beam 2 intersected it in the glass sample (Fig. 2). Polarization of the writing beams was either perpendicular ('s' polarization) or parallel ('p' polarization) to the plane defined by the sample normal and the wavevectors of the incident beams. The geometry used in our experiment determines the grating period of $\Lambda = 2.2 \mu\text{m}$. We performed a degenerate four-wave-mixing (DFWM) experiment to measure the total grating amplitude, which includes the contributions of photorefractive and other competing gratings. In the DFWM experiment, a weak probe beam 3, counterpropagating to beam 2 and proceeding from the same laser as the writing beams, is diffracted by the grating; correspondingly, beam 4 appears counterpropagating to beam 1 (Fig. 2). The grating amplitude was calculated from the ratio between the intensities of beams 4 and 3 using Kogelnik's coupled-waves theory¹⁰. A polarized He-Ne laser (wavelength 632.8 nm, output power 30 mW) was used in all experiments.

The 2BC and DFWM experiments were first performed on a poled sample tilted at $\delta = 45^\circ$ with respect to the bisector of the writing beams. Tilted geometry is required for a non-zero effective electro-optic coefficient along the grating vector $\mathbf{K}$, given the symmetry of the electro-optic tensor in films with the poling axis aligned along the film normal¹¹. The photorefractive nature of the grating was confirmed by asymmetric energy coupling between two writing beams in the 2BC experiment. The signal beam 1 was amplified and the pump beam 2 was depleted simultaneously in the glass (Fig. 3a), a consequence of the phase shift between two-beam interference fringes and the refractive index grating. A change in the direction of energy transfer, resulting from the symmetry of the electro-optic tensor in poled glasses, was observed upon rotating the sample by 180° . Using 'p' polarized writing beams, a gain of $\Gamma_p = 444 \text{ cm}^{-1}$ was measured. Given that the absorption coefficient of the film at 632.8 nm is $\alpha = 29 \text{ cm}^{-1}$, a net gain of $\Gamma_p - \alpha = 415 \text{ cm}^{-1}$ was achieved. This exceeds by far the gain of $\Gamma = 200 \text{ cm}^{-1}$ previously reported in a photorefractive sol-gel

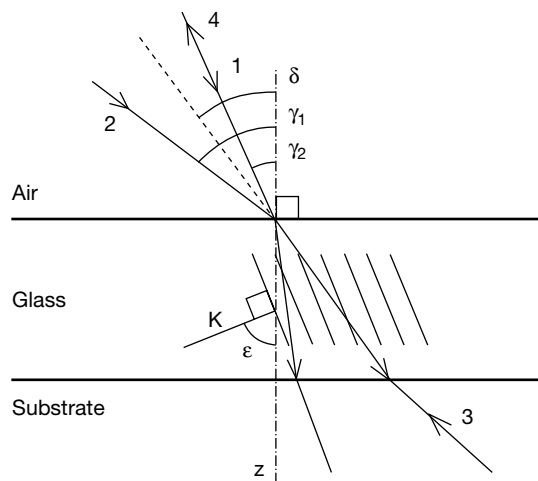


Figure 2 Geometry used for the two-beam-coupling and degenerate four-wave-mixing experiments in photorefractive glass. The tilt angle $\delta = 45^\circ$; the external angles of incidence of the writing beams 1 and 2 are $\gamma_1 = 34.5^\circ$ and $\gamma_2 = 55.5^\circ$, respectively. The bulk refractive index of the glass was measured as $n = 1.63$ by a prism-coupling technique. The calculated period of the grating is $\Lambda = 2.2 \mu\text{m}$ and the angle between the grating vector $\mathbf{K}$ and the sample normal is $\epsilon = 64.6^\circ$. The writing beams are focused to a spot of about 0.1 mm^2 and overlap in the glass. In the degenerate four-wave-mixing experiment, the readout beam 3, collinear and counterpropagating to beam 2, is focused to a spot of about 0.04 mm^2 within the grating area. The diffraction efficiency is calculated from $\eta = I_4/(I_3 - I_r)$, where I_3 and I_4 are the intensities of the read-out beam 3 and diffracted beam 4, respectively, and I_r includes Fresnel losses at the glass-substrate and glass-air interfaces. The substrate is a borosilicate glass plate coated with an indium tin oxide transparent electrode.

glass¹² and to our knowledge it is currently the highest gain reported in permanently poled photorefractive glasses and polymers. Writing with 's' polarized beams yields a gain of $\Gamma_s = 93 \text{ cm}^{-1}$. In the DFWM experiment with 's' polarized writing beams, diffraction efficiencies of $\eta_{ss} = 8.5\%$ and $\eta_{sp} = 1.1\%$ were measured for the 's' and 'p' polarized probe beams, respectively. The diffraction efficiencies for 'p' polarized writing beams were $\eta_{pp} = 25.6\%$ and $\eta_{ps} = 9.6\%$. The corresponding refractive index modulation amplitudes calculated from the coupled-wave theory are $\Delta n_{ss} = 1.9 \times 10^{-3}$, $\Delta n_{sp} = 6.8 \times 10^{-4}$, $\Delta n_{pp} = 3.5 \times 10^{-3}$ and $\Delta n_{ps} = 2.1 \times 10^{-3}$. These amplitudes include contributions from all the refractive index gratings present in our glass, both in-phase and phase-shifted with respect to the illumination intensity pattern. The amplitude of the photorefractive (phase-shifted) grating was calculated from the measured gain Γ , according to $\Delta n^{\text{PR}} = m\Lambda/(4\pi \sin \vartheta)$, where m is the modulation depth of the intensity pattern and ϑ is the phase shift between the intensity pattern and the space charge field¹³. In the absence of an external electric field the phase shift of $\vartheta = 90^\circ$ is implied from Poisson's equation¹⁴. The calculated amplitudes of the photorefractive grating for 's' and 'p' polarized writing beams are $\Delta n_s^{\text{PR}} = 4.5 \times 10^{-4}$ and $\Delta n_p^{\text{PR}} = 2.2 \times 10^{-3}$, respectively. The grating writing time was found to be similar to that observed in other permanently poled sol-gel glasses¹², being in the range of 3–5 minutes, depending on the power density of the writing beams.

For the geometry used in our experiment, the theory for permanently poled glassy polymers with electro-optic properties¹⁵ predicts the anisotropy ratio of diffraction efficiencies of the grating, when it is probed by 'p' and 's' polarized beams, to be $\eta_p/\eta_s = 8.5$. This ratio

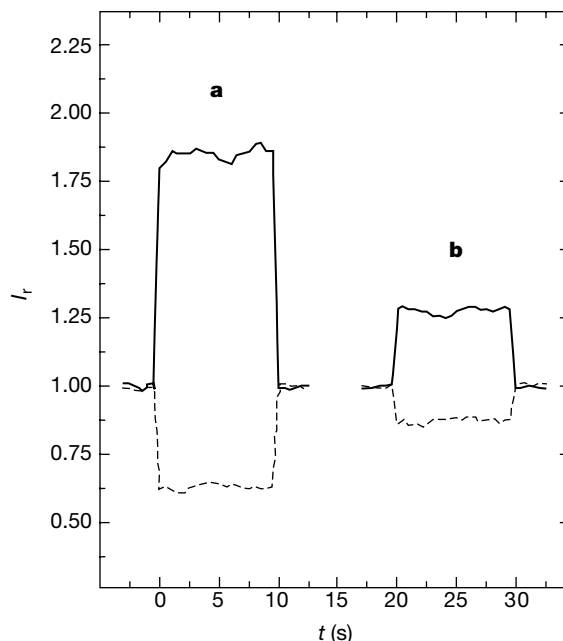


Figure 3 The two-beam-coupling signal measured in non-centrosymmetric and centrosymmetric glassy films. Relative intensity $I_t = I/I_0$ of one of the writing beams is monitored when the other beam is switched on and off at $t = t_0$ (in seconds) and $t = t_1$, respectively. Beam 1, signal beam (solid line); beam 2, pump beam (dashed line). I and I_0 are the intensities of a writing beam when the other writing beam is present or absent, respectively. Both beams are 'p' polarized and the power density of the pump beam is 3 W cm^{-2} . The gain Γ is calculated from the experimental data using $\Gamma = \{\cos \gamma_1 [\ln(r\beta/(r+1-\beta))]\}/L$, where $r = [I_2(I_1 = 0)]/[I_1(I_2 = 0)]$ is the ratio between the intensities of the writing beams 2 and 1 measured before they enter the sample; $\beta = (I_1(I_2 \neq 0))/(I_1(I_2 = 0))$ is the amplification factor; L is the film thickness; and γ_1 is the incidence angle of the beam 1 inside the film. **a**, Non-centrosymmetric glass, $t_0 = 0 \text{ s}$, $t_1 = 10 \text{ s}$, $r = 1.6$, $\beta = 1.85$ and $L = 29 \mu\text{m}$. **b**, Centrosymmetric glass, $t_0 = 20 \text{ s}$, $t_1 = 30 \text{ s}$, $r = 2$, $\beta = 1.3$ and $L = 22 \mu\text{m}$.

is constant for gratings recorded with both beams 's' polarized and both beams 'p' polarized. The diffraction efficiencies we measured in the DFWM experiment yield an anisotropy ratio for 'p' polarized writing beams of $\eta_{pp}/\eta_{ps} = 2.66$, which is 20 times higher than the ratio for 's' polarized writing beams of $\eta_{sp}/\eta_{ss} = 0.13$. This anisotropy differs significantly from that predicted by the photorefractive theory based on the electro-optic effect, but it is consistent with that previously reported for orientational non-photorefractive gratings in azo-dye composites^{16–17}.

To investigate the mechanism responsible for the photorefractive response in our glass, we carried out three separate experiments. First, 2BC and DFWM experiments were performed on an unpoled film. Theory predicts a zero gain coefficient Γ for the 2BC experiment, given that the effective electro-optic coefficient vanishes in unpoled, centrosymmetric glasses. We measured the gain coefficient of $\Gamma = 188 \text{ cm}^{-1}$ in an unpoled sample, using 'p' polarized writing beams (Fig. 3b). To our knowledge this is the first report of asymmetric two-beam coupling in a centrosymmetric material. The centrosymmetric structure of the unpoled glass was confirmed in an independent second harmonic generation experiment. This experiment eliminates the possibility of self-alignment of the chromophore during preparation of the sample, which could induce non-centrosymmetry in the glass. In the DFWM experiment on the centrosymmetric sample, we found that the diffraction efficiency for the probe beam with polarization identical to that of the writing beams, was more than 20 times higher than for the probe beam polarized perpendicularly to the writing beams.

Second, polarization recording was performed on a centrosymmetric sample according to the geometry in Fig. 2, with orthogonal polarization of the writing beams 1 and 2 and 's' or 'p' polarized read-out beam 3. In polarization recording there is no periodic variation in light intensity over the sample, and thus no periodic space-charge field can arise. Instead, there is a periodic spatial variation in the polarization state of light, which can induce orientational redistribution of dye molecules, resulting in a refractive index grating⁹. The existence of such a grating in our glass was confirmed by the diffraction of read-out beam 3 in the DFWM experiment. The non-local grating recording mechanism was not observed in the polarization recording experiment. The experiments with both polarization and intensity recording confirmed that the dye-orientational mechanism is present in our glass and that the space charge field is required for a nonlocal response in the centrosymmetric sample. The presence of the dye-orientational mechanism was also confirmed in an independent birefringence measurement, performed according to a technique reported elsewhere⁷. A steady-state birefringence of $\Delta n = 0.05$ was measured.

In a third approach, a 2BC experiment was performed on a poled sample in untitled geometry (Fig. 2, $\delta = 0$). According to photorefractive theory based on the electro-optic effect, $\Gamma = 0$ is expected for this geometry¹⁸. We measured a gain of $\Gamma = 193 \text{ cm}^{-1}$ in this experiment. This result, together with the data from the 2BC experiment in the centrosymmetric glass, confirms the nonlocal nature of the recorded grating even in the absence of the electro-optic effect.

The asymmetric two-beam coupling effect that we report in the geometries with zero effective electro-optic coefficient cannot be explained by current photorefractive theory. However, according to our observations, it is consistent with interaction between a space-charge field and a light-induced orientational grating. Given that the electric-field-induced disturbance in the dye-orientational distribution mimics the space-charge field modulation, it necessarily results in a nonlocal grating and thus in a non-zero 2BC gain, which we observed. Light-induced orientational effects in azo-dyes have been studied by several groups^{19–21}. It has been suggested and experimentally observed that in the presence of polarized light, rod-like azo molecules are pumped from the *trans* to the *cis* state

with a probability proportional to $\cos^2(\theta)$, where θ is the angle between the light polarization direction and the molecular axis^{17–19}. Given this angular dependence, the *trans* molecules that became, after their thermal relaxation from the less energetically favoured *cis* state, oriented perpendicular to the polarization of the pump beam, are extremely unlikely to be pumped back to the *cis* state. After several *trans*–*cis*–*trans* cycles, the concentration of the molecules oriented at $\theta = 90^\circ$ is progressively increased and an orientational grating is induced. Although the exact mechanism of coupling between the space-charge field and the orientational grating is not yet known, it is plausible that either photoassisted poling²² or electric-field assisted isomerization²³ may act as the coupling mechanism.

This work demonstrates recording of nonlocal, photorefractive gratings in both centrosymmetric and non-centrosymmetric organically modified silica glasses. The photorefractive properties of the glass may be enhanced by the interaction between the space-charge field and the azo-dye-orientational mechanism. The unprecedented optical gain and refractive index modulation that we achieved in our material is relevant to applications in holographic storage, coherent image amplification, self-pump phase conjugation, optical filters and limiters, and simulation of neural networks and associative memories. This multi-component glass has shown excellent resistance against chromophore crystallization and high stability of electric-field-induced chromophore alignment over a period of six months at 20 °C. The observation of a nonlocal grating in geometries with zero effective electro-optic coefficient suggests that a new class of photorefractive materials can be developed, in which non-centrosymmetric geometry or a nonlinear optical property of the chromophore are not necessarily required for a photorefractive response. □

Received 5 June; accepted 30 August 2000.

- Petrov, M. P., Stepanov, S. I. & Khomenko, A. V. *Photorefractive Crystals in Coherent Optical Systems* (Springer, Berlin, 1991).
- Nolte, D. D. (ed.) *Photorefractive Effects and Materials* (Kluwer, Dordrecht, 1995).
- Southgate, P. D. Room-temperature poling and morphology changes in pyroelectric polyvinylidene fluoride. *Appl. Phys. Lett.* **28**, 250–252 (1976).
- Brinker, C. J. & Scherer, G. W. *Sol-Gel Science. The Physics and Chemistry of Sol-Gel Processing* (Academic, San Diego, 1990).
- Ebelmen, J. J. Sur les éthers siliciques. *C.R. Acad. Sci.* **19**, 398–400 (1844).
- Meerholz, K., Volodin, B. L., Sandalphon, Kippelen, B. & Pyghambarian, N. A photorefractive polymer with high optical gain and diffraction efficiency near 100%. *Nature* **371**, 497–500 (1994).
- Chaput, E., Riehl, D., Lévy, Y. & Boilot, J.-P. New nonlinear sol-gel films exhibiting photorefractivity. *Chem. Mater.* **5**, 589–590 (1993).
- Mardner, S. R., Kippelen, B., Jen, A. K.-Y. & Peyghambarian, N. Design and synthesis of chromophores and polymers for electro-optic and photorefractive applications. *Nature* **388**, 845–851 (1997).
- Eichler, H. J., Günter, P. & Pohl, D. W. *Laser-Induced Dynamic Gratings* (Springer, Berlin, 1986).
- Kogelnik, H. Coupled wave theory for thick hologram gratings. *Bell Syst. Tech. J.* **48**, 2909–2947 (1969).
- Ducharme, S., Scott, J. C., Twieg, R. J. & Moerner, W. E. Observation of the photorefractive effect in a polymer. *Phys. Rev. Lett.* **66**, 1846–1849 (1991).
- Darracq, B. et al. Stable photorefractive memory effect in sol-gel materials. *Appl. Phys. Lett.* **70**, 292–294 (1997).
- Moerner, W. E., Grunnet-Jepsen, A. & Thompson, C. L. Photorefractive polymers. *Annu. Rev. Mater. Sci.* **27**, 586–623 (1997).
- Boyd, R. W. *Nonlinear Optics* 413 (Academic, London, 1992).
- Moerner, W. E., Silence, S. M., Hache, F. & Björklund, G. C. Orientationally enhanced photorefractive effect in polymers. *J. Opt. Soc. Am. B* **11**, 320–330 (1994).
- Mitchell, G. R., O'Leary, S. V. & Smith, M. A. Multiple grating formation in photorefractive polymers. *Polym. Preprints* **38**, 510–511 (1997).
- Sandalphon et al. Dual-grating formation through photorefractivity and photoisomerization in azo-dye-doped polymers. *Opt. Lett.* **19**, 68–70 (1994).
- Kippelen, B., Meerholz, K. & Peyghambarian, N. in *Nonlinear Optics of Organic Molecules and Polymers* (eds Nalwa, H. S. & Miyata, S.) 465–484 (CRC Press, Boca Raton, 1997).
- Todorov, T., Nikolova, L. & Tomova, N. Polarization holography. I: A new high-efficiency organic material with reversible photoinduced birefringence. *Appl. Opt.* **23**, 4309–4312 (1984).
- Gibbons, W. M., Shannon, P. J., Sun, S.-T. & Swetlin, B. J. Surface-mediated alignment of nematic liquid crystals with polarized laser light. *Nature* **351**, 49–50 (1991).
- Sekkat, Z. & Dumont, M. Polarization effects in photoisomerization of azo dyes in polymeric films. *Appl. Phys. B* **53**, 121–123 (1991).
- Sekkat, Z. & Dumont, M. Photoassisted poling of azo dye doped polymeric films at room temperature. *Appl. Phys. B* **54**, 486–489 (1992).
- Enomoto, T. et al. Electrostatically induced isomerization of azobenzene derivatives in Langmuir-Blodgett films. *J. Phys. Chem. B* **101**, 7422–7427 (1997).

Acknowledgements

We thank N. Kukhtarev, S. Janz, J. Roovers and I. Lévesque for stimulating discussions.

Correspondence and requests for materials should be addressed to F.D.M. (e-mail: delmonte@icmm.csic.es) or P.C. (e-mail: pavel.cheben@nrc.ca).

A nanometre-scale electronic switch consisting of a metal cluster and redox-addressable groups

David I. Gittins, Donald Bethell, David J. Schiffrin & Richard J. Nichols

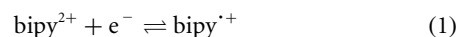
Centre for Nanoscale Science, Department of Chemistry, University of Liverpool, Liverpool, L69 7ZD, UK

So-called bottom-up fabrication methods aim to assemble and integrate molecular components exhibiting specific functions into electronic devices that are orders of magnitude smaller than can be fabricated by lithographic techniques. Fundamental to the success of the bottom-up approach is the ability to control electron transport across molecular components. Organic molecules containing redox centres—chemical species whose oxidation number, and hence electronic structure, can be changed reversibly—support resonant tunnelling^{1,2} and display promising functional behaviour when sandwiched as molecular layers between electrical contacts^{3,4}, but their integration into more complex assemblies remains challenging. For this reason, functionalized metal nanoparticles have attracted much interest^{5–7}: they exhibit single-electron characteristics^{8–10} (such as quantized capacitance charging) and can be organized^{11–13} through simple self-assembly methods into well ordered structures, with the nanoparticles at controlled locations. Here we report scanning tunnelling microscopy measurements showing that organic molecules containing redox centres can be used to attach metal nanoparticles to electrode surfaces and so control the electron transport between them. Our system consists of gold nanoclusters

a few nanometres across and functionalized with polymethylene chains that carry a central, reversibly reducible bipyridinium moiety^{14,15}. We expect that the ability to electronically contact metal nanoparticles via redox-active molecules, and to alter profoundly their tunnelling properties by charge injection into these molecules, can form the basis for a range of nanoscale electronic switches.

The approach followed in the present work is summarized schematically in Fig. 1. The redox group ('redox gate') forms the backbone of a bifunctional ligand that attaches a gold nanoparticle to a gold substrate (see Methods). Electron (or hole) injection to the redox centre is controlled by the potential applied between the substrate and a counter electrode, and the resulting change in barrier transparency can be used to control electron tunnelling between the nanoparticle and the substrate. The tip of a scanning tunnelling microscope (STM) is positioned above the nanoparticle, and is used to monitor electron tunnelling (see Methods).

The redox switchable group used here is a bipyridinium (viologen) moiety located at the centre of a chain of 20 methylene units in an α,ω -polymethylenedithiol compound (1 in Fig. 1 inset), previously described in refs 14 and 15. Reduction of the bipyridinium group takes place readily and reversibly in a self-assembled film of the ligand on a gold electrode when Au nanoparticles have been attached to the film^{14,15}; the redox process corresponds to the formation of the bipyridinium radical ion:



where bipy^{2+} represents the bipyridinium group and $\text{bipy}^{+\cdot}$ the radical.

Scanning tunnelling spectroscopy (STS)¹⁶, previously used to study electron transport through individual molecules¹⁷ and atoms¹⁸, is used to monitor the tunnelling properties of the assembly shown in Fig. 1. In STS, barrier properties are probed by altering the total potential difference between the substrate and the tip (V_{bias}) and measuring the resulting currents I (see Methods). The STS measurements, which require a constant substrate–tip distance while sweeping the bias voltage and the absence of electrochemical reactions, were conducted in mesitylene by first imaging the area of interest, then placing the tip above the chosen nanoparticle, disabling the feedback loop and immediately measuring the I – V_{bias} response. Performing this two-terminal measurement in a

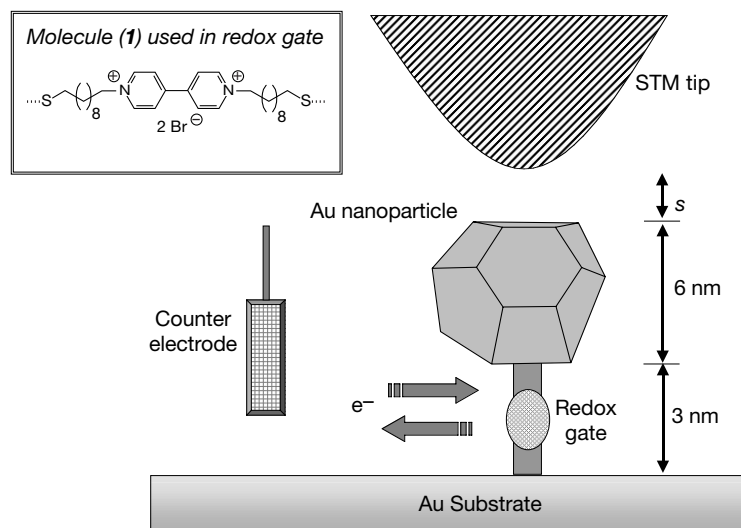


Figure 1 Schematic representation of the nanoscopic device. STM was used to examine the electrical characteristics of this device. Electrons can be injected into the redox gate by applying a suitable potential between the substrate and the counter electrode. The potential between the tip and the substrate (V_{bias}) is controlled independently, and s

represents the tip–nanoparticle distance. The redox gate consists of up to 60 of the molecules shown in the inset. The counter electrode is immersed in the surrounding electrolyte.

condensed medium and using a non-polar organic solvent allows the tip–substrate potential difference to be swept over a wide range without interference from electrochemical reactions either on the tip and substrate, or on the redox moiety present in the ligand. In addition, the structures are protected from adventitious contamination.

Figure 2 shows an STM image of isolated nanoparticles attached to the bifunctional ligand layer. Figure 3 shows the potential dependence of the current, plotted as (dI/dV_{bias}) in order to highlight the potential dependence of the tunnelling conductivity. The solid line shows the tunnelling conductance of the nanoparticle assembly containing the viologen moieties. A very sharp resonance peak can be seen for tunnelling through the structure. The dashed line corresponds to the control experiment, in which gold nanoparticles were placed over a hexanedithiol monolayer. No peak is seen in the control experiment, showing that the resonance is related to the presence of the redox ligand within the nanoparticle–metal junction.

Due to the stability of the bipy^{++} state, tunnelling through the junction is likely to occur through a temporary population of the molecular redox state in a two-step electron transfer mechanism. The STS results (Fig. 3) show that the structures investigated display a very marked and sharp resonance with a half-width of 0.2 eV, which favours such a two-step mechanism¹⁹. But the STS experiments in mesitylene do not probe changes in barrier characteristics caused by external electron injection, as happens at potentials where the redox group can accept electrons. The advantage of using redox linkers to support nanoparticles is the independent control that can

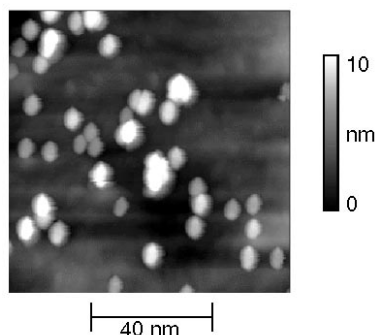


Figure 2 An STM image of a low coverage of gold nanoparticles deposited on top of a self-assembled monolayer of the redox gate molecule shown in Fig. 1. Current–voltage characteristics (STS), as well as barrier height measurements, were performed by positioning the STM tip over an isolated gold nanoparticle.

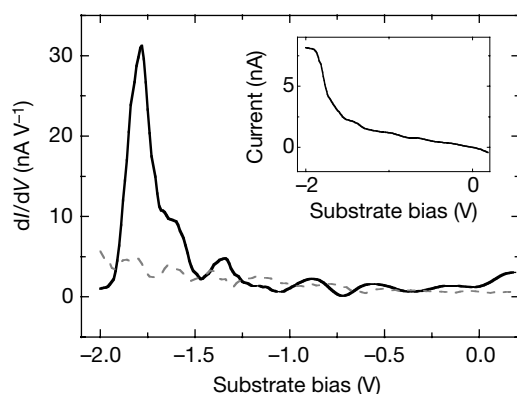


Figure 3 A current–voltage curve (inset) and its derivative (dI/dV_{bias}) obtained for the surface shown in Fig. 2, immersed in mesitylene. The STM tip was positioned above an isolated gold nanoparticle for these measurements. The dotted line corresponds to the control experiment, in which gold nanoparticles were deposited on a hexanedithiol monolayer.

be exercised over charge injection to the redox moiety (controlled with respect to a reference electrode in the solution) and the total bias applied between the particle and the substrate.

Measurements of inverse decay length¹⁶ were used to study the influence of reduction of the bipyridinium group on junction properties. These measurements were conducted in aqueous electrolyte, enabling control of the redox state of the bipyridinium group. After positioning the tip above a chosen nanoparticle, the feedback loop was switched off, the tip was advanced 0.2 nm towards the surface and the current measured as the tip was withdrawn back to its original position at a rate of 2.2 nm s^{-1} . From the average of several $I(s)$ measurements (where s is the distance between the tip and the particle), the inverse decay length was calculated from the distance dependence of $d(\ln I)/ds$ according to¹⁶:

$$\frac{d \ln I}{ds} = -2\kappa \quad (2)$$

where 2κ is the inverse decay length of the tunnelling current. κ is a measure of the decay of the electronic wavefunction away from the sample surface, as probed by the STM tip.

The inverse decay length drops sharply on reduction of bipy^{2+} to bipy^{++} , from $16 \pm 1 \text{ nm}^{-1}$ for bipy^{2+} to $7 \pm 1 \text{ nm}^{-1}$ for bipy^{++} . Previous electrochemical STS experiments²⁰ on Au(111) surfaces (also replicated by us, results not shown) indicate that the inverse decay length does not alter markedly with electrode potential: the changes observed here are due to the injection of charge into the redox group, and not to changes in the electrical double layer. Our measurements of the inverse decay length reflect the changes in the electronic properties of the junction brought about by redox switching. When the redox group in the junction is reduced, the electronic wavefunction of the gold surface, the bipy^{++} moiety and the nanoparticle extensively couple, leading to conductivity in the junction. We propose that the dramatic change in the nature of the barrier when the bipyridinium group is reduced gives rise to the large change observed in the tunnelling parameters. These observations are currently being considered within the theoretical framework of the mechanism of electron tunnelling (J. Ulstrup, personal communication).

In Fig. 4 we show the operation of a molecular switch based on the results we report here. By controlling the substrate-to-solution potential, the bipyridinium group can be switched between its oxidized (bipy^{2+}) and reduced (bipy^{++} and bipy^0) states. With the bipyridinium in its bipy^{++} reduced state (Fig. 4, centre panel), the device shows a high barrier transparency resulting from the

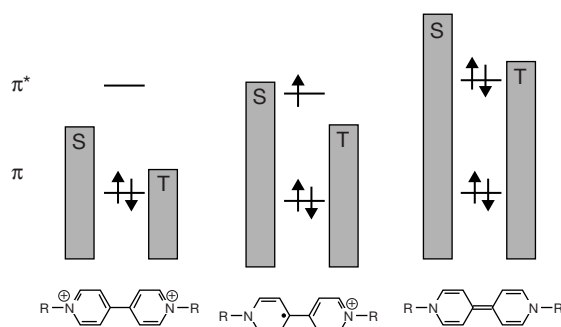


Figure 4 A schematic illustration of the operation of a nanoscopic electronic switch illustrated in Fig. 1. The figure shows the energy levels corresponding to the conduction bands of the substrate and STM tip (S, gold substrate; T, STM tip), and the lowest unoccupied molecular orbital and highest occupied molecular orbital of the bipyridinium moiety. Three redox states are shown: bipy^{2+} (left), bipy^{++} (middle) and bipy^0 (right). When the redox gate molecule is in its bipy^{++} reduced state (centre), the device shows a high barrier transparency.

extension of the electronic wavefunction from the gold surface through the bipy⁺⁺ group and the nanoparticle. The resulting increase in conductivity provides the basis for the operation of this low-voltage switching structure. The bipyridinium group has a footprint of approximately 0.45 nm² (ref. 21), and a maximum of 60 redox ligands could be attached to each gold nanoparticle. The actual number is most likely to be much less than this owing to steric hindrance by the chains, and we expect that the switch requires less than 30 electrons to operate (with half of the redox ligands in their reduced state). □

Methods

Preparation of nanostructures

The preparation of *N,N*-di-(10-mercaptodecyl)-4,4'-bipyridinium dibromide (1 in Fig. 1)¹⁴ and solutions of gold nanoparticles have been described previously^{22,23}. The gold on glass substrates were cleaned in 70:30 H₂SO₄:H₂O₂ (after use this solution must not be stored as explosive oxo-peroxides may form), and annealed using a high-purity hydrogen/air flame to yield Au(111) surfaces. Au film electrodes were derivatized by immersion for 18 h in 1 mM methanol solutions of (1)²⁴, washed with methanol and toluene, and then immersed in a toluene solution of gold nanoparticles (average size 6 nm) for a short time so as to attach well-separated particles on the surface.

STM/STS measurements

These were performed with a Pico2000 system (Molecular Imaging) using PicoScan software. STM tips (0.25 mm diameter 80:20 Pt-Ir wire, Goodfellow) were mechanically sharpened on a polishing wheel using fine abrasive paper. The STS measurements were performed using uncoated Pt-Ir tips in mesitylene, and two-probe measurements were made by altering the potential difference between the substrate and tip (*V*_{bias}), as described in ref. 2. In contrast, the inverse decay length measurements were recorded in aqueous electrolytes (0.1 M potassium phosphate buffer, pH 7) under N₂, and *V*_{bias} and the electrode potential were controlled independently using a bipotentiostat. For these experiments, the Pt-Ir STM tips were freshly coated with a layer of Apiezon wax before use to decrease leakage currents (these were measured at a tip bias of 200 mV and only tips with a leakage current lower than 5 pA were used). The microscope was operated in a constant-current mode with a tunnelling current of 50 pA and 200 mV bias (tip positive). Scans in both directions (away and towards the substrate) were recorded to ensure that the tip-sample interactions were not perturbing the results. All solutions were de-aerated with high purity nitrogen for 20 min before use.

Received 6 April; accepted 24 August 2000.

1. Tao, N. J. Probing potential-tuned resonant tunneling through redox molecules with scanning tunneling microscopy. *Phys. Rev. Lett.* **76**, 4066–4069 (1996).
2. Han, W. H. *et al.* STM contrast, electron-transfer chemistry, and conduction in molecules. *J. Phys. Chem. B* **101**, 10719–10725 (1997).
3. Feldheim, D. L. & Keating, C. D. Self-assembly of single electron transistors and related devices. *Chem. Soc. Rev.* **27**, 1–12 (1998).
4. Chen, J., Reed, M. A., Rawlett, A. M. & Tour, J. M. Large on-off ratios and negative differential resistance in a molecular electronic device. *Science* **286**, 1550–1552 (1999).
5. Ashoori, R. C. Electrons in artificial atoms. *Nature* **379**, 413–419 (1996).
6. Collier, C. P., Saykally, R. J., Shiang, J. J., Henrichs, S. E. & Heath, J. R. Reversible tuning of silver quantum dot monolayers through the metal-insulator transition. *Science* **277**, 1978–1981 (1997).
7. Oosterkamp, T. H. *et al.* Microwave spectroscopy of a quantum-dot molecule. *Nature* **395**, 873–876 (1998).
8. Chen, S. W. & Murray, R. W. Electrochemical quantized capacitance charging of surface ensembles of gold nanoparticles. *J. Phys. Chem. B* **103**, 9996–10000 (1999).
9. Ingram, R. S. *et al.* 28 kDa alkanethiolate protected Au clusters give analogous solution electrochemistry and STM coulomb staircases. *J. Am. Chem. Soc.* **119**, 9279–9280 (1997).
10. Andres, R. P. *et al.* "Coulomb staircase" at room temperature in a self-assembled molecular nanostructure. *Science* **272**, 1323–1325 (1996).
11. Kiely, C. J., Fink, J., Brust, M., Bethell, D. & Schiffrin, D. J. Spontaneous ordering of bimodal ensembles of nanoscopic gold clusters. *Nature* **396**, 444–446 (1998).
12. Schmid, G. *Clusters and Colloids* (VCH, Weinheim, 1994).
13. Templeton, A. C., Wuefeling, M. P. & Murray, R. W. Monolayer protected cluster molecules. *Acc. Chem. Res.* **33**, 27–36 (2000).
14. Gittins, D. L., Bethell, D., Nichols, R. J. & Schiffrin, D. J. Diode-like electron transfer across nanostructured films containing a redox ligand. *J. Mater. Chem.* **10**, 79–83 (2000).
15. Gittins, D. L., Bethell, D., Nichols, R. J. & Schiffrin, D. J. Redox-connected multilayers of discrete gold particles: A novel electroactive nanomaterial. *Adv. Mater.* **11**, 737–740 (1999).
16. Chen, C. J. *Introduction to Scanning Tunneling Microscopy* (Oxford Univ. Press, New York, 1993).
17. Joachim, C., Gimzewski, J. K., Schlittler, R. R. & Chavy, C. Electronic transparency of a single C₆₀ molecule. *Phys. Rev. Lett.* **74**, 2102–2105 (1995).
18. Yazdani, A., Eigler, D. M. & Lang, N. D. Off-resonance conduction through atomic wires. *Science* **272**, 1921–1924 (1996).
19. Friis, E. P. *et al.* An approach to long-range electron transfer mechanisms in metalloproteins: In situ scanning tunneling microscopy with submolecular resolution. *Proc. Natl Acad. Sci. USA* **96**, 1379–1384 (1999).
20. Vaught, A., Jing, T. W. & Lindsay, S. M. Non-exponential tunneling in water near an electrode. *Chem. Phys. Lett.* **236**, 306–310 (1995).
21. DeLong, H. C. & Buttry, D. A. Environmental effects on redox potentials of viologen groups embedded in an electroactive SAM. *Langmuir* **8**, 2491–2496 (1992).

22. Brust, M., Bethell, D., Schiffrin, D. J. & Kiely, C. J. Novel gold-dithiol nano-networks with nonmetallic electronic properties. *Adv. Mater.* **7**, 795–797 (1995).
23. Bethell, D., Brust, M., Schiffrin, D. J. & Kiely, C. From monolayers to nanostructured materials – an organic chemist's view of self-assembly. *J. Electroanal. Chem.* **409**, 137–143 (1996).
24. Ulman, A. *An Introduction to Ultrathin Organic Films* (Academic, San Diego, 1991).

Acknowledgements

This work was supported by the UK EPSRC Scanning Probe Microscopy initiative. D.I.G. thanks EPSRC for the award of a quota studentship. D.B. thanks the Leverhume Trust for the award of an Emeritus Fellowship.

Correspondence and requests for materials should be addressed to R.J.N. (e-mail: nichols@liv.ac.uk).

Large disparity between gallium and antimony self-diffusion in gallium antimonide

H. Bracht^{*,†}, S. P. Nicols^{*}, W. Walukiewicz^{*}, J. P. Silveira[‡], F. Briones[‡] & E. E. Haller^{*}

^{*} University of California and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[‡] Instituto de Microelectrónica de Madrid, Centro Nacional de Microelectrónica, CSIC, Serrano 144, 28066 Madrid, Spain

The most fundamental mass transport process in solids is self-diffusion. The motion of host-lattice ('self-') atoms in solids is mediated by point defects such as vacancies or interstitial atoms, whose formation and migration enthalpies determine the kinetics of this thermally activated process^{1,2}. Self-diffusion studies also contribute to the understanding of the diffusion of impurities, and a quantitative understanding of self- and foreign-atom diffusion in semiconductors is central to the development of advanced electronic devices. In the past few years, self-diffusion studies have been performed successfully with isotopically controlled semiconductor heterostructures of germanium³, silicon⁴, gallium arsenide^{5,6} and gallium phosphide⁷. Self-diffusion studies with isotopically controlled GaAs and GaP have been restricted to Ga self-diffusion, as only Ga has two stable isotopes, ⁶⁹Ga and ⁷¹Ga. Here we report self-diffusion studies with an isotopically controlled multilayer structure of crystalline GaSb. Two stable isotopes exist for both Ga and Sb, allowing the simultaneous study of diffusion on both sublattices. Our experiments show that near the melting temperature, Ga diffuses more rapidly than Sb by over three orders of magnitude. This surprisingly large difference in atomic mobility requires a physical explanation going beyond standard diffusion models. Combining our data for Ga and Sb diffusion with related results for foreign-atom diffusion in GaSb (refs 8, 9), we conclude that the unusually slow Sb diffusion in GaSb is a consequence of reactions between defects on the Ga and Sb sublattices, which suppress the defects that are required for Sb diffusion.

Compared to conventional self-diffusion experiments, which use radioactive probe atoms of the matrix introduced from the surface, diffusion studies with highly enriched stable isotopes are neither limited by the half-life of the radiotracer nor affected by near-surface degradation. Such diffusion experiments can therefore be extended to much longer times and conducted over a wider temperature range. In combination, these advantages lead to more accurate data for the activation enthalpy of self-diffusion.

[†]Present address: University of Münster, Institut für Materialphysik, Wilhelm-Klemm-Strasse 10, D-48149 Münster, Germany.

Here we use an isotope structure consisting of alternating GaSb layers with different isotopic compositions to investigate the diffusion of both Ga and Sb in GaSb.

Figure 1a shows concentration profiles of the stable Ga and Sb isotopes of an undoped $^{69}\text{Ga}^{121}\text{Sb}/^{71}\text{Ga}^{123}\text{Sb}$ isotope heterostructure measured using secondary ion mass spectrometry (SIMS). The structure was grown by molecular beam epitaxy (MBE) at 450 °C on a 200-nm-thick undoped natural GaSb buffer layer that was deposited on a (100)-oriented GaSb substrate wafer. Specially designed valved effusion cells were used for both the Ga and the Sb isotopes. Raman spectroscopy of the epitaxial layer revealed a sharp phonon line spectrum that is typical of undoped GaSb bulk crystals, and shows the high crystalline quality of the layers grown by MBE. Samples, $1 \times 2 \text{ mm}^2$ in size, were cut from the wafers, and placed (with the epitaxial layers face-up) on a polished 500- μm -thick undoped GaSb wafer with lateral dimensions of $9 \times 10 \text{ mm}^2$; these assemblies were placed in diffusion ampoules. Self-diffusion experiments at temperatures between 571 °C and 708 °C under

Ga-rich (Sb-rich) conditions were performed by adding $(30 \pm 3) \text{ mg}$ of $\text{Ga}_{0.65}\text{Sb}_{0.35}$ ($\text{Ga}_{0.40}\text{Sb}_{0.60}$) into the diffusion ampoules, which had volumes of typically $(2.0 \pm 0.3) \text{ cm}^3$. The diffusion annealing cycles were performed in a resistance-heated tube furnace; the temperature was monitored with an accuracy of $\pm 2 \text{ K}$, and a correction of 2–3 min for heating-up time was allowed for when determining the total time of diffusion. The data in Fig. 1b show that annealing of the isotope structure at 700 °C for 105 min causes a pronounced intermixing of the Ga isotope layers but no detectable change of the corresponding Sb structure. After annealing for about 18 d at 700 °C, a small intermixing of the Sb isotope layers could be observed (Fig. 1c) while the Ga isotopes had reached a homogeneous distribution.

The temperature dependences of the Ga and the Sb diffusion coefficients in GaSb are shown in Fig. 2. The results for Ga diffusion under Sb- and Ga-rich conditions are identical. The close agreement with the diffusion of the isovalent impurity indium in GaSb under Ga-rich conditions indicates that the composition of the as-grown isotope sample belongs to the Ga-rich branch of the GaSb phase diagram. Accordingly the Ga diffusion data are representative for Ga-rich conditions. Attempts to measure Sb diffusion for Ga-rich conditions failed because of insufficient Sb motion. The Sb diffusion data shown in Fig. 2 represent the results for Sb-rich conditions that were established during the Sb diffusion experiments at least as quickly as the time it takes for the Ga isotopes to intermix. The Sb diffusion coefficient remains extremely small, close to $10^{-18} \text{ cm}^2 \text{ s}^{-1}$, even near the melting point of GaSb (712 °C). The large disparity between the diffusion coefficients of the components is very unusual, considering that within the zinc blende structure each group-III atom is surrounded by four group-V atoms and vice versa. To emphasize the extent of the observed asymmetry of the self-diffusion on the two sublattices, we point out that it takes more than one hour for an Sb atom to diffuse a distance corresponding to the

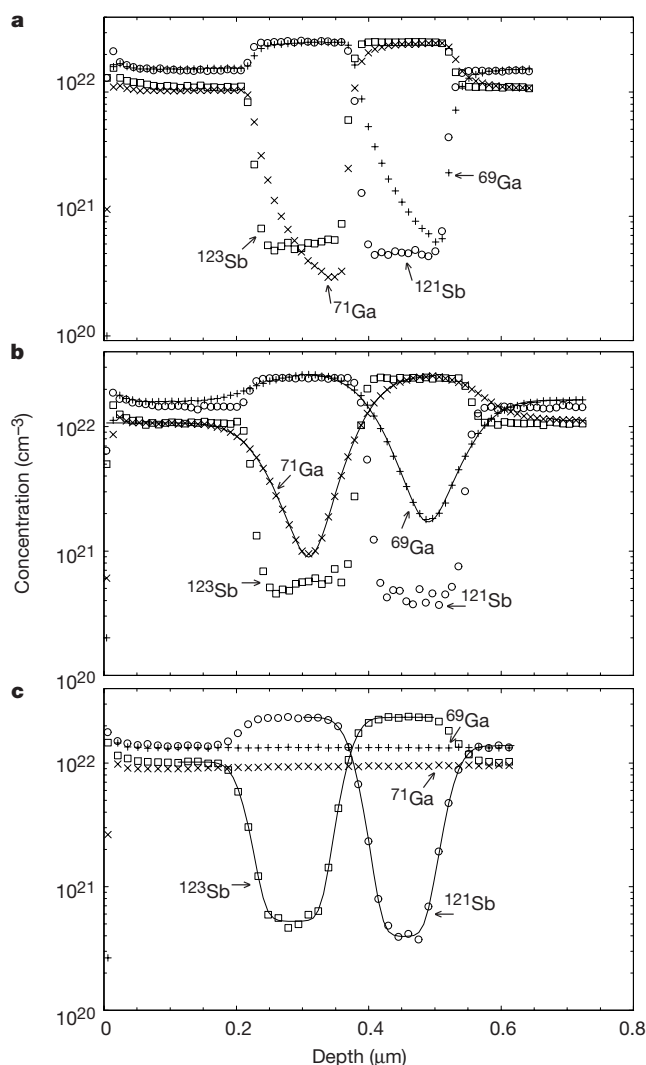


Figure 1 Concentration–depth profiles of Ga and Sb isotopes in GaSb isotope heterostructures measured with SIMS. Data are shown for ^{69}Ga (plus symbols), ^{71}Ga (crosses), ^{121}Sb (circles) and ^{123}Sb (squares). **a**, Ga and Sb profiles of the as-grown $^{69}\text{Ga}^{121}\text{Sb}/^{71}\text{Ga}^{123}\text{Sb}$ heterostructure. **b,c**, Profiles after annealing the isotope structure under Sb-rich conditions at 700 °C for 105 min (**b**) and 18 d (**c**). For clarity, only every second data point is plotted. The Ga and Sb diffusion coefficients were determined by fitting the solution of Fick's law for self-diffusion across an interface⁶ to the Ga and Sb profiles measured after annealing. Solid lines in **b** and **c** show best fits.

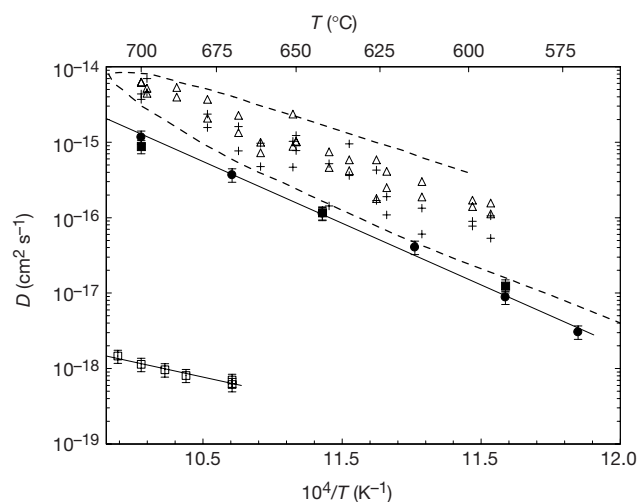


Figure 2 Temperature dependence of the diffusion coefficients D of Ga, Sb and In in GaSb. The figure shows our measurements of the self-diffusion coefficient D for Ga in Sb-rich GaSb (circles) and in Ga-rich GaSb (filled squares); also shown are our measurements of D for Sb in Sb-rich GaSb (open squares). Significant diffusion of Sb under Ga-rich conditions was not observed. We estimate an upper limit of $1.5 \times 10^{-18} \text{ cm}^2 \text{ s}^{-1}$ for the Sb diffusion coefficient at 700 °C under Ga-rich conditions. Our diffusion data are accurate to within 20%. This accuracy mainly stems from the uncertainty in measuring (with a surface profilometer) the depth of the craters formed by SIMS. The lower (upper) solid line represents the result of fitting an Arrhenius equation $D = D_0 \exp(-Q/kT)$ to the Sb (Ga) diffusion data for Sb-rich conditions, which yields $\ln(D_0/\text{cm}^2 \text{ s}^{-1}) = -22.3 \pm 2.0$ (4.41 ± 1.27) and $Q/\text{eV} = 1.59 \pm 0.17$ (3.24 ± 0.10). The diffusion data from ref. 13 for Ga (triangles) and Sb (plus signs) in Sb-rich conditions are shown for comparison. The lower and upper dashed curves illustrate the temperature dependence of indium diffusion in Ga- and Sb-rich GaSb, respectively⁸.

GaSb lattice parameter (0.61 nm): Ga atoms travel this distance in a few seconds.

Similar disparities in self-diffusion rates of compounds have been seen in transition-metal silicides such as Fe₃Si (ref. 10) and FeSi (ref. 11). But in these systems the diffusion coefficients of both elements approach each other near the melting point, with values in the range of 10^{-10} cm² s⁻¹ and higher, which are typical for solids crystallizing in a cubic B20 structure (FeSi) and a body-centred structure (Fe₃Si) (ref. 1). Compared with self-diffusion in pure metals and metallic compounds, self-diffusion in elemental semiconductors Si and Ge near the melting point is characterized by lower values of about 10^{-12} cm² s⁻¹, a reflection of the strong covalent bonding^{2,4}. Moreover, the diffusion of Ga and As in GaAs, which have been determined by isotope heterostructures^{5,6} and radiotracer diffusion experiments¹², respectively, yielded self-diffusion coefficients at the melting point of about 3×10^{-13} cm² s⁻¹ and 5×10^{-14} cm² s⁻¹. These, and related results from other material systems investigated so far, considerably exceed the diffusivity measured for Sb in GaSb near its melting point.

The large difference between Ga and Sb diffusion in GaSb suggests that both self-atoms diffuse independently on their own sublattices. Earlier investigations of GaSb led to models with coupled diffusion mechanisms, which involve a native defect formed by two Ga vacancies (V_{Ga}) and a Ga anti-site atom (Ga_{Sb}) (ref. 13). Our results contradict this so-called triple defect mechanism. A possible explanation for the discrepancy between the earlier results and ours may be a degradation of the near-surface region. Such a degradation can not be excluded, even though additional GaSb powder was added during the diffusion anneals (H. Mehrer, personal communication). In fact, using transmission electron microscopy (TEM), we have observed that natural GaSb samples annealed under Ga-rich conditions show a degradation of the near-surface region extending to about 50 nm into the bulk. The previously measured radioactive Sb profiles originating at the sample surface extend to such depths. Presumably, degradation-induced defects are the cause for the discrepancy in the Sb diffusion data. Such degradation could also be the cause for the larger scatter in the earlier diffusion data (Fig. 2). But the deeply buried isotope heterostructures that we have used are much less susceptible to such degradation.

The diffusion coefficient of each component of GaSb is given by a contribution of vacancies (V_{Ga} , V_{Sb}) and interstitials (Ga_i , Sb_i) on the respective sublattices:

$$D_{Ga,Sb} = C_{V_{Ga,Sb}}^{eq} D_{V_{Ga,Sb}} + C_{Ga_i,Sb_i}^{eq} D_{Ga_i,Sb_i} \quad (1)$$

Here $C_{V_{Ga,Sb}}^{eq}$ and C_{Ga_i,Sb_i}^{eq} are the thermal equilibrium concentrations of vacancies and self-interstitials, and $D_{V_{Ga,Sb}}$ and D_{Ga_i,Sb_i} are the individual diffusion coefficients. We note that this equation neglects the contribution of the anti-site defects to self-diffusion via a direct exchange between nearest-neighbour and second-nearest-neighbour atoms, or complex collective motion of the host atoms. Such contributions neither result in a long-range transport of the host atoms nor are they expected to be energetically more favourable than the diffusion via vacancies and self-interstitials. Collective motion of self-atoms has been considered for solids such as metallic supercooled liquids and glasses¹⁴, but not for crystalline semiconductors. Equation (1) shows that the atomic motion of self-atoms depends on the equilibrium concentrations of native defects. In the case of III–V compound semiconductors, these concentrations not only depend on temperature, but also on the partial pressure of the group-V element that becomes significant at elevated temperatures.

Extensive studies on the dependence of In diffusion in GaSb on the pressure of Sb (ref. 8), and corresponding theoretical calculations of Ga vacancy concentrations⁹, reveal that In diffuses via a vacancy mechanism on the Ga sublattice. The temperature dependence of Ga diffusion in GaSb that we report here resembles that of

the isovalent impurity In under Sb-poor conditions (see Fig. 2). This provides strong evidence that Ga also diffuses mainly via vacancies on its own sublattice, although this native defect is not expected to dominate under Ga-rich conditions. Such conditions should favour Ga interstitials (Ga_i), Ga anti-site atoms (Ga_{Sb}), and Sb vacancies (V_{Sb}). We have not observed intermixing of the Sb isotope structure, although V_{Sb} should be present. This apparent inconsistency may be removed by assuming that V_{Sb} undergoes a relaxation process, in which an adjacent Ga atom (Ga_{Ga}) moves into the Sb vacancy creating an anti-site defect and a Ga vacancy. This transformation is represented by:



Accordingly, Ga vacancies are formed even under Ga-rich conditions, and facilitate Ga diffusion. At the same time Sb vacancy formation is suppressed, and so is Sb diffusion via this native defect. The large suppression of the Sb self-diffusion may be strongly related to the charge states of the native defects. At high temperatures, used in our diffusion experiments, the GaSb samples are intrinsic; that is, the Fermi level is located slightly above the middle of the bandgap, favouring the formation of acceptor-like defects on the right-hand side of reaction (2) (refs 15–17). Consequently, the contribution of the donor-like defect V_{Sb} (refs 15, 16) as a vehicle for Sb diffusion under Ga-rich conditions is suppressed. Compared to reaction (2), the reaction



may describe the corresponding transformation under Sb-rich conditions. Reaction (3) also represents a transformation between donor-like (Sb_{Ga}) and acceptor-like (V_{Ga}) defects. The equilibrium of this reaction is considered to lie on the left-hand side, due to the Coulomb attraction between the Sb_i donor and the V_{Ga} acceptor. This, however, does not imply that the concentration of Ga vacancies is small, because the equilibrium concentration of V_{Ga} may be maintained by the supply of V_{Ga} from the surface. Once these vacancies approach Sb interstitials, the formation of an Sb anti-site is highly probable, yielding a low concentration of Sb interstitials and a correspondingly small Sb diffusion.

Our attempt to explain the small Sb diffusivity suggests that Sb interstitials mediate the diffusion on the Sb sublattice. In accord with this conclusion, the temperature dependence of Sb diffusion also points to an Sb_i -mediated process. Fitting an Arrhenius equation $D = D_0 \exp(-Q/kT)$ to the temperature dependence of the Sb diffusion data, we obtain a pre-exponential factor, $\ln(D_0/\text{cm}^2 \text{s}^{-1})$, of -22.3 ± 2.0 , and an activation enthalpy, Q , of (1.59 ± 0.17) eV. The low value for D_0 indicates that the concentration of the native defect mediating Sb diffusion must increase with decreasing temperature, otherwise unphysically large negative values for the entropy of formation and migration of the native defect would result. Of the two candidates (V_{Sb} and Sb_i) that could mediate the diffusion of Sb on its own sublattice under Sb-rich conditions, only Sb_i could exhibit a concentration that increases with decreasing temperature.

In contrast to earlier experimental results, our investigations of Ga and Sb diffusion in GaSb with isotope heterostructures reveal that Ga and Sb diffuse independently on their own sublattices. The observed unusually slow diffusion of Sb in GaSb can be explained by transformations among native point defects. Under Ga-rich conditions (see equation (2)), the Sb vacancy is unstable and can not contribute to Sb self-diffusion whereas Ga vacancies can be easily formed and mediate the self-diffusion of Ga on the Ga sublattice. The transformation under Sb-rich conditions (see equation (3)) implies small concentrations of Sb interstitials, and thus explains the extremely slow interstitial-mediated diffusion of Sb on the Sb sublattice. Analogous transformations between native defects have been proposed to explain the doping limitations of III–V compound semiconductors as well as the stabilization of the Fermi level

observed after high-energy irradiation of these materials^{18,19}. The present diffusion study provides additional evidence for this physical concept, and illustrates the strong correlation between atomic mass transport processes and electronic properties of native defects in semiconductors. □

Received 6 July; accepted 11 September 2000.

1. Landolt-Börnstein, New Series, Group III, Vol. 26, Pt B, *Diffusion in Solid Metals and Alloys* (ed. Mehrer, H.) (Springer, Berlin, 1990).
2. Landolt-Börnstein, New Series, Group III, Vol. 33, Pt A *Diffusion in Semiconductors* (ed. Beke, D. L.) (Springer, Berlin, 1998).
3. Fuchs, H. D. et al. ⁷⁰Ge/⁷⁴Ge isotope heterostructures: an approach to self-diffusion studies. *Phys. Rev. B* **51**, 16817–16821 (1995).
4. Bracht, H., Haller, E. E. & Clark-Phelps, R. Silicon self-diffusion in isotope heterostructures. *Phys. Rev. Lett.* **81**, 393–396 (1998).
5. Wang, L. et al. Ga self-diffusion in GaAs isotope heterostructures. *Phys. Rev. Lett.* **76**, 2342–2345 (1996).
6. Bracht, H., Norseng, M., Haller, E. E., Eberl, K. & Cardona, M. Enhanced and retarded Ga self-diffusion in Si and Be doped GaAs isotope heterostructures. *Solid State Commun.* **112**, 301–314 (1999).
7. Wang, L. et al. Gallium self-diffusion in gallium phosphide. *Appl. Phys. Lett.* **70**, 1831–1833 (1997).
8. Mathiot, D. & Edelin, G. Diffusion of indium in GaSb. *Phil. Mag. A* **41**, 447–458 (1980).
9. Edelin, G. & Mathiot, D. A model for the determination of the defect concentrations in III-V compounds: The case of GaSb. *Phil. Mag. A* **42**, 95–110 (1980).
10. Gude, A. & Mehrer, H. Diffusion in the D0₃-type intermetallic phase Fe₃Si. *Phil. Mag. A* **76**, 1–29 (1997).
11. Salamon, M. & Mehrer, H. Diffusion in the B20-type phase FeSi. *Phil. Mag. A* **79**, 2137–2155 (1999).
12. Bösker, G. et al. Arsenic diffusion in intrinsic gallium arsenide. *Mater. Res. Soc. Symp. Proc.* **527**, 347–356 (1998).
13. Weiler, D. & Mehrer, H. Self-diffusion of gallium and antimony in GaSb. *Phil. Mag. A* **49**, 309–325 (1984).
14. Tang, X.-P., Geyer, U., Busch, R., Johnson, W. L. & Wu, Y. Diffusion mechanisms in metallic supercooled liquids and glasses. *Nature* **402**, 160–162 (1999).
15. Talwar, D. N. & Ting, C. S. Tight-binding calculations for the electronic structure of isolated vacancies and impurities in III-V compound semiconductors. *Phys. Rev. B* **25**, 2660–2680 (1982).
16. Xu, H. Electronic structure of neutral and charged vacancies in Ga-related III-V compound semiconductors. *J. Appl. Phys.* **68**, 4077–4086 (1990).
17. Kühn, W., Strehlow, R. & Hanke, M. Self-consistent tight-binding investigation of chemical trends for native defects in III-V semiconductors. *Phys. Status Solidi B* **151**, 541–557 (1987).
18. Walukiewicz, W. Mechanism of Fermi level stabilization in semiconductors. *Phys. Rev. B* **37**, 4760–4763 (1988).
19. Walukiewicz, W. Amphoteric native defects in semiconductors. *Appl. Phys. Lett.* **54**, 2094–2096 (1989).

Acknowledgements

H.B. acknowledges a Feodor Lynen fellowship from the Alexander von Humboldt-Stiftung. This work was supported in part by the Office of Energy Research, Office of Basic Energy Sciences, Materials Science Division of the US Department of Energy, and the US National Science Foundation.

Correspondence and requests for materials should be addressed to H.B. (e-mail: bracht@uni-muenster.de) or to E.E.H. (e-mail: eehaller@lbl.gov).

Early onset and tropical forcing of 100,000-year Pleistocene glacial cycles

Scott Rutherford & Steven D'Hondt

University of Rhode Island, Graduate School of Oceanography, Narragansett, Rhode Island 02882, USA

Between 1.5 and 0.6 Myr ago, the period of the Earth's glacial cycles changed from 41 kyr, the period of the Earth's obliquity cycles, to 100 kyr, the period of the Earth's orbital eccentricity^{1,2}, which has a much smaller effect on global insolation. The timing of this transition and its causes pose one of the most perplexing problems in palaeoclimate research³. Here we use complex demodulation to examine the phase evolution of precession and semiprecession cycles—the latter of which are phase-coupled to both precession and eccentricity—in the tropical and extra-

tropical Atlantic Ocean. We find that about 1.5 Myr ago, tropical semiprecession cycles (with periods of about 11.5 kyr) started to propagate to higher latitudes, coincident with a growing amplitude envelope of the 100-kyr cycles. Evidence from numerical models suggests that cycles of about 10 kyr in length may be required to explain the high amplitude of the 100-kyr cycles⁴. Combining our results with consideration of a modern analogue, we conclude that increased heat flow across the equator or from the tropics to higher latitudes around 1.5 Myr ago strengthened the semiprecession cycle in the Northern Hemisphere, and triggered the transition to sustained 100-kyr glacial cycles.

The 100-kyr band of global ice-volume variability exhibits several significant features over the past 2.6 Myr (Fig. 1a). What are generally regarded as the late Pleistocene glacial cycles began about 0.6 Myr ago and the first high-amplitude 100-kyr variability began about 1.2 Myr ago⁵. Inspection of the filtered record shows that the amplitude envelope of the 100-kyr cycle began to expand toward late Pleistocene values 1.4–1.5 Myr ago. Before then, 100-kyr cycles occurred but were not sustained.

This interpretation is supported by an elegant study of the amplitude and phase of the $\delta^{18}\text{O}$ record at two major eccentricity periods (95 and 124 kyr)⁶. The study showed that frequency modulation of $\delta^{18}\text{O}$ in the eccentricity band, which is characteristic of the late Pleistocene⁷, and the phase lock between $\delta^{18}\text{O}$ and orbital eccentricity began near 1.5 Myr ago⁶. Furthermore, although its temporal resolution is limited, another study showed that the asymmetry (a third-moment quantity that describes the shape of a time series) of the $\delta^{18}\text{O}$ record changed dramatically 1.5 Myr ago⁸.

The timing of this transition coincided with other events

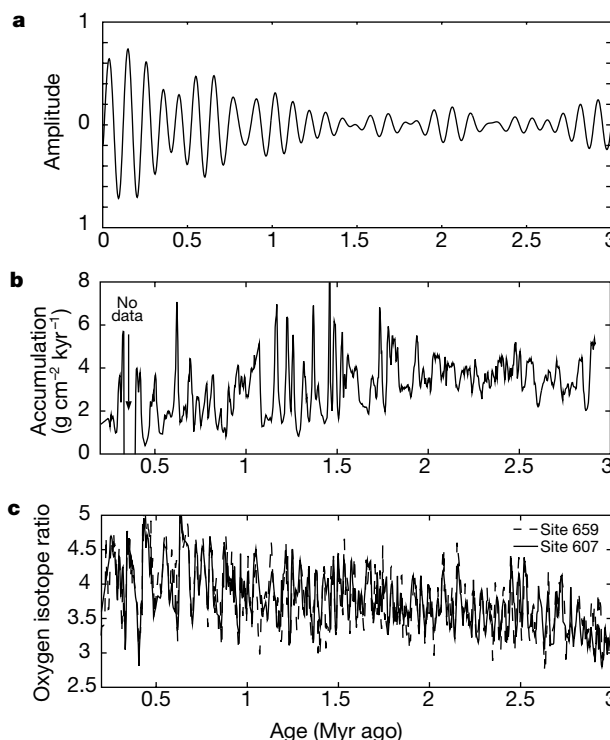


Figure 1 Oxygen isotope and calcium carbonate accumulation time series. **a**, Bandpass filter of the Site 659 $\delta^{18}\text{O}$ record. Periods between 90 and 130 kyr are passed by the filter. We note that the band begins to oscillate consistently and increase in amplitude near 1.5 Myr ago. **b**, Calcium carbonate mass accumulation for Site 607. Mass accumulation is calculated from weight per cent carbonate^{27,28}, wet bulk density measured by the Gamma Ray Attenuation Porosity Evaluator, grain density³⁰ and sedimentation rate calculated from the oxygen isotope age model. We note the change in the character of CaCO_3 accumulation that occurs near 1.5 Myr ago. **c**, $\delta^{18}\text{O}$ record from Site 607 correlated to $\delta^{18}\text{O}$ from Site 659.

observed in the palaeoclimate record. At about 1.6 Myr ago, glacial–interglacial $\delta^{13}\text{C}$ variations and dissolution of CaCO_3 in the tropical Atlantic indicate the first substantial decrease in North Atlantic Deep Water (NADW) production during glacial periods^{9,10}. Palaeoclimate records of the Asian monsoon exhibit major phase changes relative to insolation¹¹ and indices of African aridity indicate a change to more arid conditions at this time¹². In addition, the rate of loess deposition in China increased and the loess/palaeosol alternations began to correlate with the ice-volume record ($r = 0.7$ from 0.0–1.5 Myr ago and $r = 0.15$ from 1.5–2.5 Myr ago) at 1.5 Myr ago¹³. Although sustained high-amplitude glacial cycles did not begin until about 1.2 Myr ago⁵, we suggest the transition to a climate regime conducive to sustained 100-kyr ice-sheet variability occurred 1.5 Myr ago.

One explanation for the high-amplitude climate response to low-amplitude orbital forcing is that the 100-kyr glacial cycles are a free oscillation internal to the climate system with the phase set by Milankovitch forcing^{14,15}. If so, ice-sheet nonlinearities would contribute to a climate system that is inherently sensitive to forcing at eccentricity periods. However, this hypothesis does not adequately explain why 100-kyr climate fluctuations exist in a number of pre-Pleistocene climate records when large ice sheets were not available to provide the 100-kyr sensitivity, or when ice sheets were oscillating at other frequencies^{16,17}.

Some numerical models also require an approximately 10-kyr free oscillation or time constant that is internal to the model and dependent on free parameters^{4,14,15}. For example, when a model⁴ was run without an internal ~ 10 -kyr oscillation the 100-kyr response to weak eccentricity forcing was correspondingly small. Subsequent adjustment of the free parameters in the model produced an internal, unforced ~ 10 -kyr cycle. Model runs with the internal ~ 10 -kyr oscillation and the weak eccentricity forcing produced a much larger 100-kyr response to weak eccentricity forcing. This result is consistent with resonance of a nonlinear oscillator to additive forcing⁴. Another model demonstrated a strong linkage between low-latitude semiprecession and 100-kyr cycles¹⁸. In the model output, the amplitude of the eccentricity-band spectral peaks is proportional to the amplitude of the semiprecession band spectral peaks.

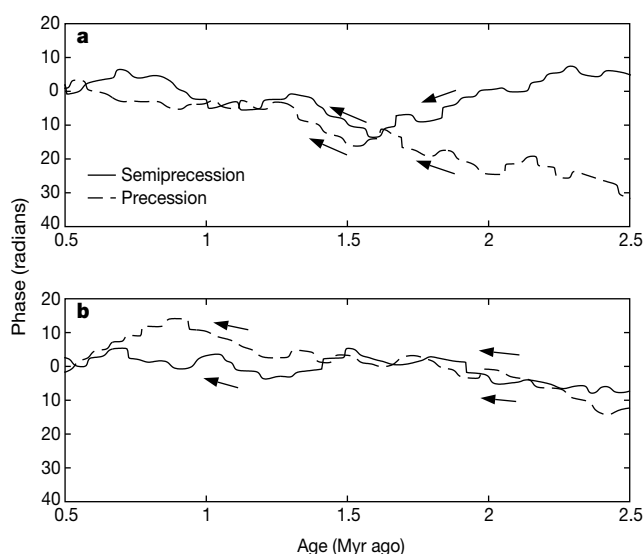


Figure 2 Instantaneous phase determined by complex demodulation. **a**, Phase of Site 607 CaCO_3 mass accumulation at precession and semiprecession frequencies. We note that before 1.5 Myr ago the phases move in opposite directions, indicating independence. After 1.5 Myr ago the phases move in the same direction, indicating they are coupled. **b**, Phase of Site 659 dust flux at precession and semiprecession frequencies. The phases are coupled throughout the entire 2.6-million-year record.

Our bispectral analyses (see Methods) suggest that semiprecession cycles, which are phase-coupled to precession and eccentricity, may provide the ~ 10 -kyr oscillation needed by these climate models to produce 100-kyr ice-volume cycles. If this is the case, we would expect that the phase coupling between precession, semiprecession and eccentricity, which originates in the tropics, propagated to the high latitudes near the onset of 100-kyr glacial cycles.

A previous study used bispectral analysis to show that phase couplings similar to those in the insolation solutions were present in ice-volume records from 0 to 1.0 Myr ago but not from 1.0 to 2.6 Myr ago⁸. To further investigate these phase couplings we examined CaCO_3 mass accumulation rate (MAR) estimates from Deep Sea Drilling Project (DSDP) North Atlantic Site 607 (41°N , 32°W) (Fig. 1b) and dust accumulation estimates from Ocean Drilling Program (ODP) tropical Atlantic Site 659 (18°N , 21°W) on the same age model (Fig. 1c) (see Methods)¹⁹. The results of our bispectral analysis of these time series suggest that the phase-coupled characteristic of the tropical semiprecession cycles propagated to the North Atlantic between 1.2 and 1.9 Myr ago (see Methods).

If two frequencies are phase-coupled, as indicated by the bispectrum, then the instantaneous phases of the two frequencies (which can be determined by complex demodulation) should change in unison over time. Complex demodulation of the Site 607 CaCO_3 MAR at precession and semiprecession periods indicates that these two frequencies were phase-coupled in the North Atlantic between 0.5 and ~ 1.5 Myr ago but not before ~ 1.5 Myr ago (Fig. 2). This result is supported by a change in the character of the CaCO_3 MAR record near 1.5–1.6 Myr ago (Fig. 1b). In contrast, complex demodulation of the Site 659 dust flux record indicates that the precession and semiprecession periods were phase-coupled in the tropical Atlantic for at least the last 2.6 Myr, confirming the bispectral results. These results suggest that phase-coupled semiprecession cycles, which originate in the tropics, propagated to the high latitudes near 1.5 Myr ago, the same time that persistent 100-kyr ice-volume cycles began to build (Fig. 1a).

There are two ways to generate semiprecession cycles that are phase-coupled to precession and eccentricity. First, the alignment of perihelion with each equinox during one precession cycle produces a semiprecession cycle, analogous to the semiannual cycle produced by the passage of the sun over the equator twice each year. An energy balance model¹⁸ demonstrated this process and its ability to produce power at 100- and 400-kyr periods. When semiprecession cycles occurred in the model output, 100-kyr cycles also occurred, with power proportional to that of the semiprecession cycle. In this case tropical heat export (either atmospheric or oceanic) can propagate the semiprecession signal from low to high latitudes¹⁸. Second, because the Northern and Southern hemispheres are 180° out of phase with respect to precession, the export of a Southern Hemisphere precession signal to the Northern Hemisphere will produce a semiprecession cycle in the Northern Hemisphere. Today, a Southern Hemisphere annual signal is exported to the Northern Hemisphere in the western tropical Atlantic where the south equatorial current is diverted into the Northern Hemisphere at Point Calcanhar, Brazil. Case two is similar to case one in that both require two precession cycles that are 180° out of phase. In case one, the cycles occur when perihelion is aligned with each equinox, whereas for case two they occur when perihelion is aligned with each solstice.

An annual analogue for both cases occurs in the tropical Atlantic. For case one, the passage of the sun over the equator twice per year produces a semiannual cycle, just as alignment of an equinox with perihelion twice during each precession cycle produces a semiprecession cycle. A semiannual cycle is documented in an empirical orthogonal function analysis of tropical Atlantic sea surface temperature (SST). In this analysis the second eigenvector exhibits

semiannual SST cycles with SST maxima occurring during both boreal spring and boreal autumn²⁰.

The export of a seasonal signal from the Southern to the Northern Hemisphere is an annual analogue for case two because the hemispheres are 180° out of phase with respect to both the seasonal and precessional cycles. Between 20° N and 20° S in the Atlantic Ocean, total monthly heat content peaks during austral summer and generally follows the Southern Hemisphere annual cycle²¹. In the northern tropical Atlantic (0° to 20° N), however, a second heat content peak occurs during boreal summer. In addition, the eastern equatorial Atlantic (6° N to 6° S) exhibits periods of heat export during both boreal summer and austral summer. Thus a semiannual cycle exists in the northern tropical Atlantic owing to the cross-equatorial influence of the Southern Hemisphere. Furthermore, tropical Atlantic SST and evaporative heat flux are important components in modelling the phase of the North Atlantic oscillation²² which in turn influences the production of NADW on interannual to decadal timescales.

There is evidence that the semiannual analogue of heat export and SST in the eastern tropical Atlantic applies to semiprecession cycles. Estimates of SST and thermocline depth (an important factor in heat storage and export) in the tropical Atlantic over the last 20 kyr show two peaks approximately 10 kyr apart²³. The double peak is particularly strong in the east, as in the semiannual analogue. In addition, zonal wind-driven divergence in the equatorial Atlantic over the past 40 kyr exhibits a dominant periodicity near that of semiprecession²⁴.

The African and Asian monsoons provide another means of producing semiprecession cycles. Previous studies showed that the strength of the Asian monsoon is in phase with a Southern Hemisphere latent heat source at the precession period¹¹ while the African monsoon is in phase with maximum Northern Hemisphere summer insolation²⁵. Neither of these studies examined the phase of semiprecession cycles, but they suggest that monsoon strength depends on sensible heating of the continent in the Northern Hemisphere and latent heat transported from the Southern Hemisphere. Northern Hemisphere sensible heating is greatest when boreal summer is aligned with perihelion and Southern Hemisphere latent heat export is greatest one-half precession cycle later, when boreal summer is aligned with aphelion and atmospheric circulation during austral winter is strongest¹¹. By shifting the balance of these monsoon drivers, it may be possible to vary the intensity of a semiprecession cycle in the Northern Hemisphere.

Recent modelling studies of orbitally-forced El Niño variability provide additional support for a tropical driver of ice-sheet variability²⁶. The model output shows that El Niño frequency and intensity varies at precession and semiprecession periods. Extending the modern El Niño effects on Hadley circulation, poleward heat transport, and high-latitude North American seasonality to longer timescales suggests that tropical teleconnections may be important in the growth and decay of large Northern Hemisphere ice sheets²⁶.

Our results indicate that a major climate threshold was crossed near 1.5 Myr ago; a time marked by large changes in low-latitude terrestrial conditions^{11,13,25} and in thermohaline circulation^{9,10}. Our results also suggest that enhancement of either low- to high-latitude climate connections (case 1) or cross-equatorial heat flow (case 2) initiated the growth of 100-kyr ice sheets at that time. In either case, our results suggest that the tropics played a major role in the initiation and maintenance of the 100-kyr ice-sheet oscillations of the last 1.5 million years. □

Methods

Data

We placed the Site 607 CaCO₃ and Site 659 dust accumulation records on the oxygen isotope^{2,27} age model from Site 659 (ref. 19). Although weight per cent CaCO₃ at Site 607 is strongly influenced by ice-rafted detritus (IRD) during the past 2.5 Myr (ref. 28), CaCO₃

MAR depends on dissolution at depth and CaCO₃ production in the surface waters and is independent of dilution by IRD. The Site 659 dust flux record is a particularly relevant tropical record because low-latitude land masses should exhibit strong semiprecession variability¹⁸.

Signal processing

Bispectral analysis is a signal processing technique that detects phase coupling between different frequencies in a single time series. Bispectral analysis of Site 607 CaCO₃ MAR shows significant North Atlantic phase couplings between precession, eccentricity, and semiprecession from 0.2 to 1.2 Myr ago but not from 1.9 to 2.9 Myr ago. However, bispectral results for ODP Site 659 (18° N, 21° W) dust flux¹⁹ indicate that significant tropical phase couplings existed during both time intervals.

Like spectral analysis, the results of bispectral analysis are averaged over the length of the time series and it is difficult to constrain the timing of transitions. To determine the timing of the transition to a phase-coupled system in the North Atlantic, we used complex demodulation to extract the instantaneous phase of the phase-coupled frequencies determined by bispectral analysis. Demodulation was performed using the ARAND package on phase-locked precession and semiprecession periods as indicated by the bispectral results. For Site 607 these were 19 and 9.5 kyr which show strong phase couplings with each other and with eccentricity over the past 1.6 million years. For Site 659 the demodulation isolated periods of 21 and 10.5 kyr. These two periods are strongly phase-coupled with each other and with eccentricity throughout the past 2.6 million years. A filter with 100 weights, a half-amplitude frequency of 0.01 cycles per kyr and a stop frequency of 0.016 cycles per kyr was used on all demodulations. The filter satisfies two requirements²⁹: it passes all frequencies out to one-half the precession (or semiprecession) bandwidth (~0.0078); and it safely stops all harmonics, which begin at a frequency of 0.028 after demodulation (for example, the first precession harmonic 1/12 kyr to 1/18 kyr).

Received 2 February; accepted 5 September 2000.

1. Raymo, M. E., Ruddiman, W. F. & Clement, B. M. in *Initial Reports Deep Sea Drilling Project: Leg 96* (eds Ruddiman, W. F. et al.) 895–901 (US Government Printing Office, Washington DC, 1986).
2. Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Backman, J. Pleistocene evolution: Northern Hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography* **4**, 353–412 (1989).
3. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1131 (1976).
4. LeTreut, H. & Ghil, M. Orbital forcing, climatic interactions, and glaciation cycles. *J. Geophys. Res.* **88**, 5167–5190 (1983).
5. Chen, J., Farrell, J. W., Murray, D. W. & Prell, W. L. Timescale and paleoceanographic implications of a 3.6 m.y. oxygen isotope record from the northeast Indian Ocean (Ocean Drilling Program site 758). *Paleoceanography* **10**, 21–47 (1995).
6. Park, J. & Maasch, K. A. Plio-Pleistocene time evolution of the 100-kyr cycle in marine paleoclimate records. *J. Geophys. Res.* **98**, 447–461 (1993).
7. Rial, J. A. Pacemaking the ice ages by frequency modulation of Earth's orbital eccentricity. *Science* **285**, 564–569 (1999).
8. Hagelberg, T., Pisias, N. & Elgar, S. Linear and nonlinear couplings between orbital forcing and the marine $\delta^{18}\text{O}$ record during the Late Neogene. *Paleoceanography* **6**, 729–746 (1991).
9. Raymo, M. E., Ruddiman, W. F., Shackleton, N. J. & Oppo, D. W. Evolution of Atlantic-Pacific $\delta^{18}\text{O}$ gradients over the last 2.5 m.y. *Earth Planet. Sci. Lett.* **97**, 353–368 (1990).
10. Bickert, T., Cordes, R. & Wefer, G. in *Proceedings of the Ocean Drilling Program, Scientific Results, Leg 154* (eds Shackleton, N. J., Curry, W. B., Richter, C. & Bralower, T. J.) 229–237 (Ocean Drilling Program, College Station, Texas, 1997).
11. Clemens, S. C., Murray, D. W. & Prell, W. L. Nonstationary phase of the Plio-Pleistocene Asian monsoon. *Science* **274**, 943–948 (1996).
12. deMenocal, P. & Bloemendal, J. in *Paleoclimate and Evolution with Emphasis on Human Origins* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 262–288 (Yale Univ. Press, New Haven, Connecticut, 1995).
13. Bloemendal, J., Liu, X. M. & Rolph, T. C. Correlation of the magnetic susceptibility of Chinese loess and the marine oxygen isotope record: Chronological and palaeoclimatic implications. *Earth Planet. Sci. Lett.* **131**, 371–380 (1995).
14. Maasch, K. A. & Saltzman, B. A low-order dynamical model of global climatic variability over the full Pleistocene. *J. Geophys. Res.* **95**, 1955–1963 (1990).
15. Saltzman, B. & Verbitsky, M. Late Pleistocene climate trajectory in the phase space of global ice, ocean state, and CO₂: Observations and theory. *Paleoceanography* **9**, 767–779 (1994).
16. Herbert, T. D. & Fischer, A. G. Milankovitch climate origin of mid-Cretaceous black shale rhythms in central Italy. *Nature* **321**, 739–743 (1986).
17. D'Hondt, S., King, J. & Gibson, C. Oscillatory marine response to the Cretaceous-Tertiary impact. *Geology* **24**, 611–614 (1996).
18. Short, D. A., Mengel, J. G., Crowley, T. J., Hyde, W. T. & North, G. R. Filtering of Milankovitch cycles by Earth's geography. *Quat. Res.* **35**, 157–173 (1991).
19. Tiedemann, R., Sarnthein, M. & Shackleton, N. J. Astronomic timescale for the Pliocene Atlantic $\delta^{18}\text{O}$ and dust flux records from Ocean Drilling Program Site 659. *Paleoceanography* **9**, 619–638 (1994).
20. Servain, J. & Legler, D. M. Empirical orthogonal function analysis of tropical Atlantic sea surface temperature and wind stress: 1964–1979. *J. Geophys. Res.* **91**, 14181–14191 (1986).
21. Merle, J. Seasonal variation of heat-storage in the tropical Atlantic Ocean. *Oceanol. Acta* **3**, 455–463 (1980).
22. Rodwell, M. J., Rowell, D. P. & Folland, C. K. Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature* **398**, 320–323 (1999).
23. Molino, B. & McIntyre, A. Nutricline variation in the equatorial Atlantic coincident with the Younger Dryas. *Paleoceanography* **5**, 977–1008 (1990).
24. McIntyre, A. & Molino, B. Forcing of Atlantic equatorial and subpolar millennial cycles by precession. *Science* **274**, 1867–1870 (1996).
25. deMenocal, P. B. Plio-Pleistocene African climate. *Science* **270**, 53–59 (1995).

26. Clement, A., Seager, R. & Cane, M. Orbital controls on the El Nino/Southern Oscillation and the tropical climate. *Paleoceanography* **14**, 441–456 (1999).
27. Raymo, M. E., Ruddiman, W. F., Backman, J., Clement, B. M. & Martinson, D. G. Late Pliocene variation in Northern Hemisphere ice sheets and North Atlantic Deep Water circulation. *Paleoceanography* **4**, 413–446 (1989).
28. Ruddiman, W. F. & Raymo, M. in *Initial Reports of the Deep Sea Drilling Project: Leg 96* (eds Ruddiman, W. F., Kidd, R. B. & Thomas, E.) 855–878 (US Government Printing Office, Washington, 1986).
29. Bloomfield, P. *Fourier Analysis of Time Series: An Introduction* 1–258 (Wiley, New York, 1976).
30. Ruddiman, W. F., Kidd, R. B., Thomas, E. et al. *Initial Reports of the Deep Sea Drilling Program: Leg 96* 1251 (US Government Printing Office, Washington DC, 1987).

Acknowledgements

We thank R. Tiedemann for providing the data from Site 659 and M. Raymo for data from Site 607. K. Maasch and M. E. Mann provided helpful comments. This work was supported by the National Science Foundation.

Correspondence and requests for materials should be addressed to S.R. (e-mail: srutherford@gso.uri.edu).

Accelerated hydration of the Earth's deep crust induced by stress perturbations

Bjørn Jamtveit, Håkon Austrheim & Anders Mølthe-Sørensen

The Fluid Rock Interaction Group, Departments of Geology & Physics, University of Oslo, PO Box 1047 Blindern, N-0316 Oslo, Norway

The metamorphic cycle associated with the formation of mountain belts produces a lower crust containing little or no free fluid^{1,2}. The introduction of external fluids to dry and impermeable volumes of the Earth's crust is thus a prerequisite for the retrogressive metamorphism later observed in such regimes. Such metamorphism can cause significant changes in the crust's physical properties, including its density, rheology and elastic properties^{3,4}. On a large scale, the introduction of fluids requires the presence of high-permeability channels, such as faults or

fractures, which are the result of external tectonic stresses. But extensive interaction between externally derived fluids and the fractured rock requires efficient mass transport away from the initial fractures into the rock itself, and this transport often occurs over distances much longer than expected from grain-boundary diffusion. Here we present both field observations and a simple network model that demonstrate how the transport of fluids into initially dry rock can be accelerated by perturbations in the local stress field caused by reactions with fluids. We also show that the morphology of reaction fronts separating 'dry' from 'wet' rocks depends on the anisotropy of the external stress field.

'Dry' volumes of the Earth's crust contain no separate fluid phase although both the mineral surfaces and crystal lattices are likely to contain more or less strongly adsorbed volatile components. Reactions between volatiles and such dry rocks may occur at any depth in the Earth's crust at which externally derived fluids get access to the rocks at a temperature where hydrous phases or carbonates are stable. An excellent field example of such a process can be found in the Bergen Arcs of southwestern Norway. The Bergen Arcs are a series of arcuate Caledonian nappes thrust upon the Proterozoic basement rocks of the western gneiss region of western Norway⁵ (Fig. 1).

One of these, the Lindaas nappe, consists mainly of dry Proterozoic granulite facies rocks, including anorthosites, that were partly hydrated during the Caledonian orogeny some 420 Myr ago⁶. The hydration occurred both under eclogite and amphibolite facies metamorphic conditions. Figure 2a shows how a reaction front defining the transition from dry anorthosite to hydrated ('wet') eclogite surrounds a fracture that initially formed within the anorthosite. This fracture is filled with a vein assemblage consisting of euhedral omphacite in a matrix of quartz, clinozoisite, phengite, plagioclase and kyanite that precipitated from the externally introduced fluid at a temperature around 700 °C and a pressure around 1.5 GPa (ref. 7).

Phase petrological analyses indicate that the fluid was water-rich (>50 mol.% H₂O)⁷. Apart from a kink fold centred on the fracture or vein (Fig. 2a), there is no extensive internal deformation within the eclogitized part of the anorthosite, and the front separating the anorthosite from the eclogite is flat and morphologically stable on the macroscale. The transition from anorthosite to eclogite involves

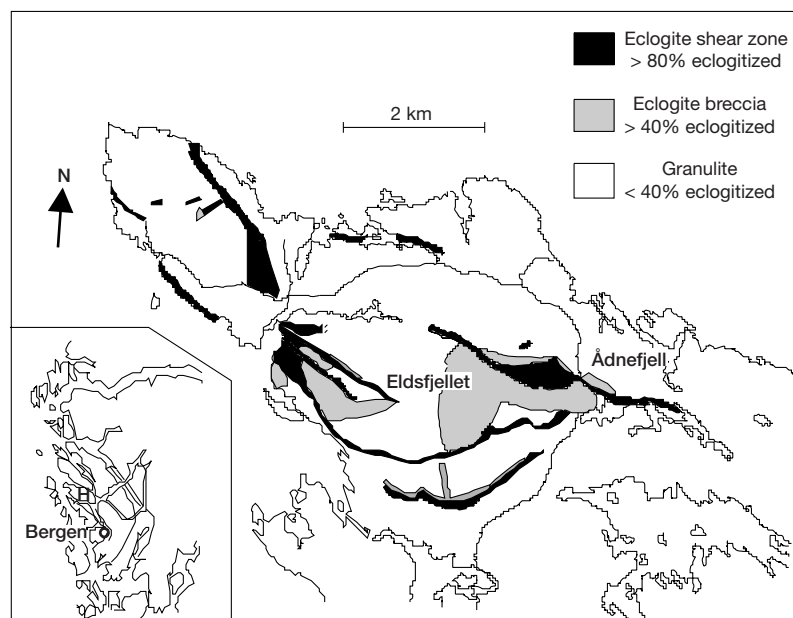


Figure 1 Geological map of Holsnøy northwest of Bergen in western Norway. Holsnøy is labelled H in the inset. Non-shaded regions are composed of 'dry' granulite facies rocks,

notably anorthosites. Shadings denote areas with different extents of eclogitization (hydration under eclogite facies conditions).

a reduction in the total solid volume by about 10 vol.% (ref. 3) and, apart from the addition of H₂O, the reaction is shown to be close to isochemical⁸. Furthermore, phase petrological studies indicate that the fluid-triggered eclogitization reactions, at least initially, took place far from thermodynamic equilibrium⁷. For a reaction overstep of 100 °C and typical values of the entropy of hydration reactions and the molar volumes of the phases involved, the corresponding isotropic stress change caused by the reaction (given as $\Delta G/V$) would be around 3 kbar (ref. 9).

Often the eclogitized part of the anorthosite develops into shear zones, reflecting the presence of an anisotropic external stress field even after brittle fracturing. It also reflects the ductile rheology of the hydrated eclogite compared to the dry anorthosite. When large volumes of the anorthosite have been transformed into eclogite, the whole volume of rock, which may crop out on a kilometre scale, becomes ductile and rigid blocks of unreacted anorthosite are

observed 'floating' around in eclogites that seem to have flowed like a viscous fluid⁴.

When external stresses cause shearing of the eclogites, the hydration front separating the anorthosite from the eclogite becomes unstable and shows fingering. Figure 2b shows two eclogitic fingers moving into dry anorthosite from an eclogitic shear zone, and Fig. 2c shows two fingers that originate in two different shear zones located on opposite sides of an anorthosite layer. In such cases, the fingers seem to stop growing when their tips reach the same distance from the shear zones. On a microscopic scale, the texture of the eclogitic rock is characterized by fine-grained symplectitic intergrowth of the eclogite facies minerals (Fig. 2d) reflecting the high overstepping of the eclogite-producing reactions (see page 203 in ref. 10 for scaling of intergrowths).

The original granulite facies garnets and pyroxenes were extensively fractured normal to the direction of finger propagation

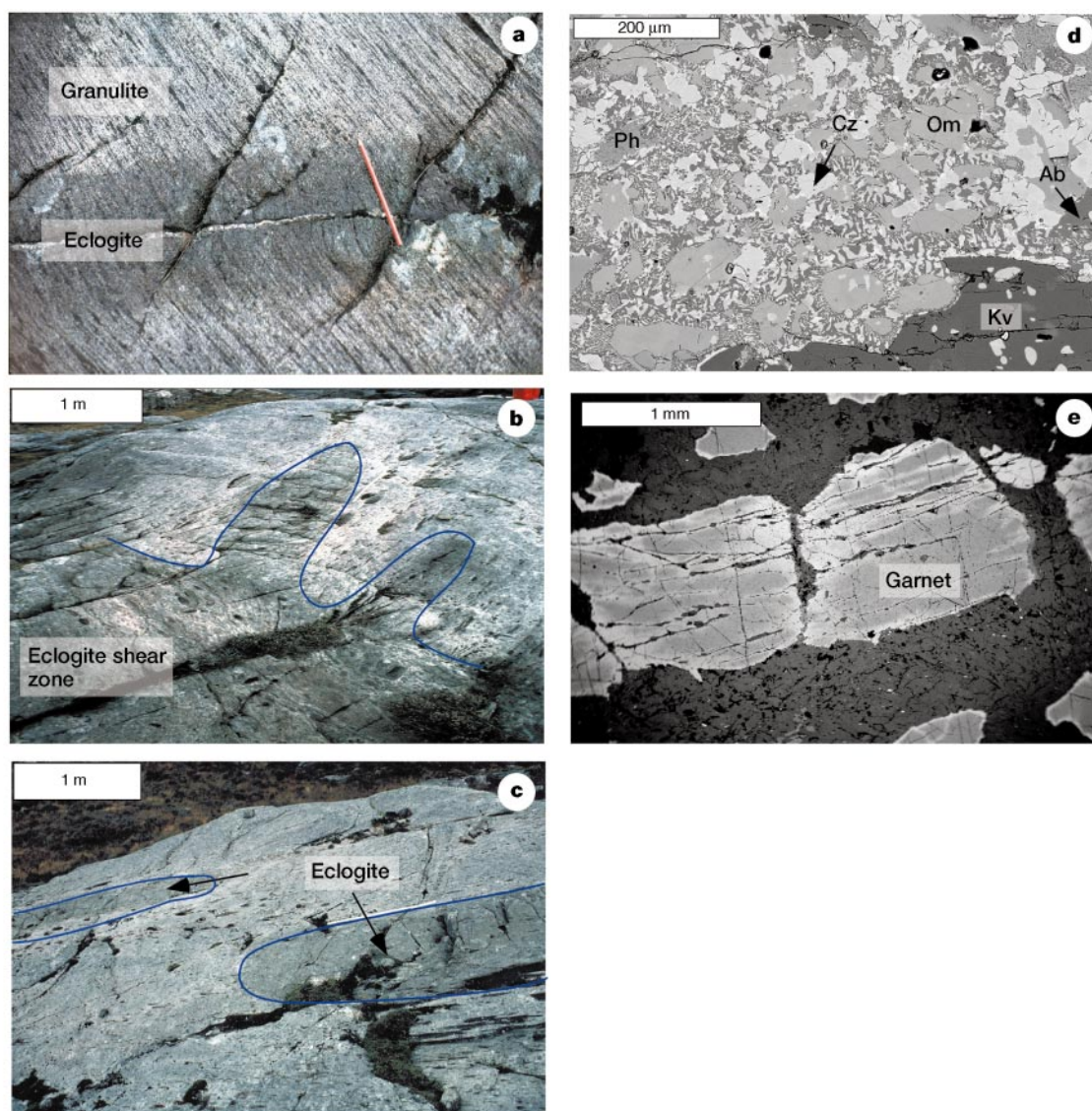


Figure 2 Field and petrographic characteristics of eclogitization fronts and eclogitized rocks. **a**, 10-cm-wide and morphologically stable eclogitization fronts around eclogite facies vein through granulite facies anorthosites. We note the limited deformation in the eclogitized part (dark colour). **b**, Eclogite fingers starting from eclogite facies shear zones (lower left corner), penetrating about two metres into anorthosites. **c**, Two eclogite fingers starting in adjacent shear zones (located just outside the edges of the picture). Fingers seem to stop growing when they reach the same distance from the shear zones. **d**, Back-scattered electron (BSE) image showing the symplectitic microtexture of an eclogitized,

initially plagioclase-rich, domain of the granulite. The fine-grained intergrowths of omphacitic pyroxene (Om), clinozoisite (Cz), phengite (Ph), albite (Ab) and kyanite (Kv) reflect short-range mass transport and are consistent with high overstepping and reaction rates during eclogite formation. **e**, BSE-image of fractures through granulite facies garnet within the eclogitized volume of the pre-existing anorthosite. Parallel cracks are partly sealed by precipitated eclogite facies minerals (omphacite, kyanite, epidote-minerals). Some cracks are healed, but are still visible because of a change in garnet composition towards higher Fe-content (lighter colour) at eclogite facies conditions.

(Fig. 2e). Some micro-fractures healed, and can only be observed as a change in the garnet composition along the pre-existing fracture; others are partly sealed by precipitation of eclogite facies minerals. However, apart from fractures in the garnets, most cracks that formed during front propagation were probably obliterated by later recrystallization, crack healing and plastic deformation.

We propose that the propagation of the eclogitization front is driven by surface fracturing induced by the phase transitions occurring at the interface between 'dry' and 'wet' rocks. When the anorthosite surface is transformed to eclogite, the associated volume reduction induces tensile stresses that eventually lead to surface fracturing. The newly generated fractures allow fluids to penetrate further into the anorthosite, where the surface again shrinks and fractures.

We have developed a dynamic model in order to understand this hydration process, estimate its propagation velocity, and address the effects of the imposed stress fields. A discrete-element approach is used^{11–13}, which allows new fractures to nucleate, and the front

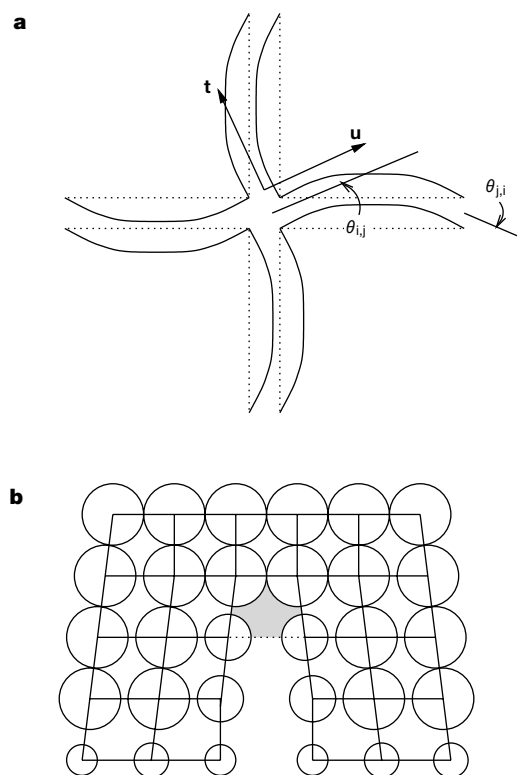


Figure 3 The simulation model. **a**, The disks that compose the elastic material are connected by elastic beams. The force and moment in each beam depend on the relative positions and angles of rotation of the disks, and are calculated using standard beam theory¹⁸ (see ref. 13 for details). The elastic properties of the elements can be directly related to the large-scale elastic coefficients. Time is increased in small steps. At each step the radius of each node along the fluid–rock interface is calculated, and a new equilibrium configuration of the elastic network is found using over-relaxation techniques¹⁹. Beams that exceed their threshold stresses are removed, and the network relaxed again, repeating the process until no more beams are broken. **b**, The node radius is reduced along the reaction front owing to the eclogitization reaction. **b** illustrates the effect of a beam rupture along the front. When the dotted beam is removed, the fluid flows through the newly generated fracture into the shaded area, and the eclogitization reaction is initiated at the surrounding nodes. The different radii along the front indicate that the eclogitization process is not uniform along the front, but depends on the exposure time since fluid contact. The fluid is assumed to be connected to a large fluid reservoir at a constant isostatic pressure. Thus, during reaction-induced volume reduction, fractures nucleate within the eclogitized volumes and propagate into the anorthosite, allowing fluids to physically flow into the initially dry rocks during front advancement.

morphology to be addressed. The elastic behaviour of the rock is modelled by a network of elastic elements, disks, connected by elastic beams (Fig. 3)¹⁴. The beams can only sustain a maximum stress, which, if exceeded, results in their irreversible removal. The behaviour of the material is determined by the distribution of breaking thresholds, which is related to the fracture toughness¹⁵.

The eclogitization reaction along the fluid–rock interface is assumed to be rapid compared to the rate of fluid transport. Consequently, the reaction zone grows into the anorthosite as a sharp diffusion front. Fracturing occurs down to the smallest scale L at which completely eclogitized fragments do not fracture¹⁶. We choose a disk size equal to L , making the process a surface reaction

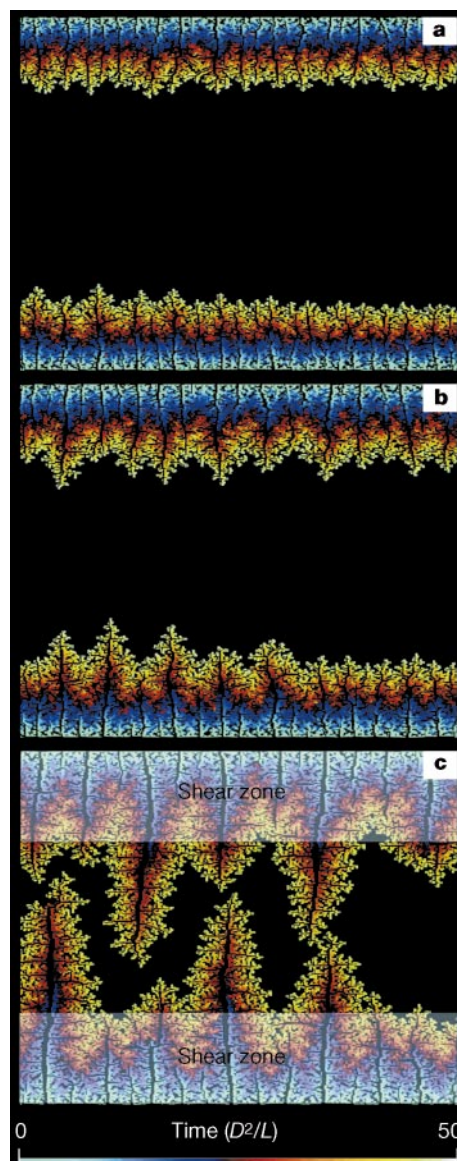


Figure 4 Simulated fracture patterns for the fluid invasion process. Initially, only the top and bottom nodes are in contact with the fluid. Nodes are coloured according to the exposure time to the fluid with a common timescale at the bottom. Time is measured in units of D^2/L , the reaction time for a single grain. The effect of increasing the anisotropy in the external stress field, that is, increasing pre-existing stress, σ_0 , is shown. **a**, $\sigma_0 = 0$; **b**, $\sigma_0 = 0.005E$; **c**, $\sigma_0 = 0.010E$, where E is Young's modulus. The fingering instability becomes more pronounced as σ_0 is increased. Even for $\sigma_0 = 0$ the front has some roughness, but here the roughness does not grow with time. The shading in **c** shows the part of the eclogitized rock that is expected to have developed into a shear zone as a response to the anisotropic stress during or after the growth of the fingers.

in our representation. As the reaction progresses, the radii of the surface elements are gradually reduced, corresponding to the local change in volume. Eventually, a surface beam ruptures, and the fluid–rock interface is extended to include the newly generated fracture surfaces (Fig. 3). This model demonstrates how the concerted action of volume-changing mineral reactions and material failure can generate tree-like connected fractures (Fig. 4).

The front propagation velocity was determined to be $v = D/L$, where D is a coefficient describing the rate of fluid migration into the dry rock in the absence of cracks. In the following, we will refer to D as a ‘bulk’ diffusion coefficient as our model does not deal with the details of the migration mechanisms involved. In a polycrystalline material, fractures predominantly occur along grain boundaries, and L corresponds to the grain size. However, as shown in Fig. 2e for large garnet grains, fracturing also occurs internally in each grain. Fig. 2e indicates a fragmentation length L_g (ref. 15) of around 0.1 mm. The width of the hydrated zone at time t is $X = vt = Dt/L$ for the front propagation mechanism described here, but only $x = (Dt)^{1/2}$ in the case of pure diffusion. Figure 5 shows that for a diffusion front advancement in excess of 0.1 mm, the fracture-propagated front will have progressed further than a diffusion front for a fragmentation length (L_g) of 0.1 mm. Thus for any timescale longer than is required to propagate a diffusion front for 0.1 mm, the mechanism proposed here is more efficient than grain boundary diffusion in transporting fluids into dry rocks. This timescale is about one month for a diffusion constant of $10^{-15} \text{ m}^2 \text{ s}^{-1}$ (ref. 17).

This model produces a morphologically flat eclogitization front. Any perturbation extending beyond the front is overtaken by the rest of the front as growth progresses. Consequently, the invasion process alone cannot explain the observed eclogite fingering (Fig. 2b and c). However, Fig. 4b and c shows that the presence of an externally imposed anisotropic stress field, which must have been present during formation of the eclogite shear zones, may induce fingering. Several propagating fingers are apparent in Fig. 4c.

Both in the field and in the model the fingers stop growing when they meet fingers from adjacent shear zones. This is a result of local stress relief. The width of the individual fingers is comparable to the thickness of eclogitized layers for morphologically stable fronts, but the propagation lengths of the fingers are considerably longer. Thus,

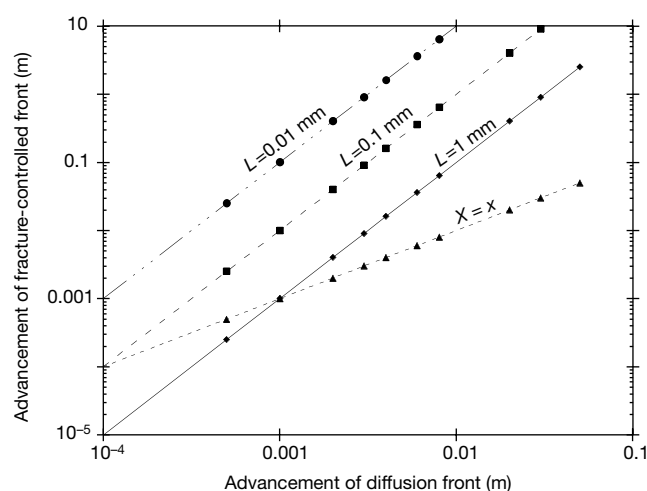


Figure 5 Comparison of fracture-propagated versus purely diffusion-controlled hydration fronts. Shown is the advancement distance of the eclogite–granulite front for the fracture-controlled propagation process we propose here (X) relative to the advancement distance of a pure diffusion front (x) if both have grown over equally long time. The curve $X = x$ shows the situation when both have grown equally far from the fluid inlet. For $L = 0.1$ mm, $X > x$ if $x > 0.1$ mm. If, for example, the time available for transport has been sufficient to advance a diffusion front by 1 mm (about 3 years for $D > 10^{-15} \text{ m}^2 \text{ s}^{-1}$), the fracture front would have advanced more than an order of magnitude further.

the presence of even a modestly anisotropic external stress field effectively speeds up the rate of pervasive fracturing and fluid infiltration into the dry rocks through its coupling with local stress perturbations caused by mineral reactions.

The close correspondence between the field observations and our simple elastic spring network model, suggest that this model captures the most important parts of the dynamics controlling the advancement of the eclogitization front. Mineral reactions seem both to cause changes in rock rheology and to perturb the local stress field in the area around the hydration front. Rocks and externally introduced fluids, in most cases, will be far from equilibrium at the onset of the metamorphic reactions; this ensures that the rate of the metamorphic reactions is not rapidly slowed down or buffered by the stress changes or thermal effects of the reaction itself.

The mechanism of fluid introduction to dry rocks proposed here is particularly efficient in the common case where tectonic or other external forces generate an anisotropic stress field in the system where reactions with volatiles occur. Extensive and relatively large-scale fracturing with associated fluid-facilitated mass transport may occur even though the externally imposed deviatoric stresses alone are insufficient to cause fracturing. We believe this mechanism is the key to understanding how dry crustal volumes may undergo retrogressive metamorphism on a scale that is several orders of magnitude longer than would be expected from an interaction mechanism only controlled by grain boundary diffusion. □

Received 2 March; accepted 12 September 2000.

- Yardley, B. W. D. in *Fluid Flow and Transport in Rocks: Mechanisms and Effects* (eds Jamveit, B. & Yardley, B. W. D.) 99–121 (Chapman and Hall, London, 1997).
- Yardley, B. W. D. & Valley, J. W. The petrologic case for a dry lower crust. *J. Geophys. Res.* **102**, 12173–12185 (1997).
- Austrheim, H. Eclogitization of lower crustal granulites by fluid migration through shear zones. *Earth Planet. Sci. Lett.* **81**, 221–232 (1987).
- Austrheim, H. Eclogite formation and the dynamics of crustal roots under continental collision zones. *Terra Nova* **3**, 492–499 (1991).
- Sturt, B. A., Skarpenes, O., Ohanian, A. T. & Pringle, T. R. Reconnaissance Rb/Sr isochron study in the Bergen Arc System and regional implications *Nature* **253**, 595–599 (1975).
- Bingen, B., Davis, W. J. & Austrheim, H. Zircon growth during fluid induced Caledonian/Scandian eclogite-facies metamorphism of the Lindås Nappe, Caledonides of W Norway. *Min. Mag.* **62**, 161–162 (1998).
- Jamveit, B., Bucher-Nurminen, K. & Austrheim, H. Fluid controlled eclogitization of granulites in deep crustal shear zones, Bergen Arcs, western Norway. *Contrib. Mineral. Petrol.* **104**, 184–193 (1990).
- Rockow, K. M., Haskin, L. A., Jolliff, B. L. & Fountain, D. M. Constraints on element mobility associated with the conversion of granulite to eclogite along fractures in an anorthositic complex on Holsnoy, Norway. *J. Metamorph. Geol.* **15**, 401–418 (1997).
- Wheeler, J. The significance of grain-scale kinetics of metamorphism. *Contrib. Mineral. Petrol.* **97**, 397–404 (1987).
- Kirkaldy, J. S. in *Growth, Dissolution and Pattern Formation in Geosystems* (eds Jamveit, B. & Meakin, P.) 189–220 (Kluwer, Dordrecht, 1999).
- Cundall, P. A. & Strack, O. D. L. A discrete numerical model for granular assemblies. *Geotechnique* **29**, 47–65 (1979).
- de Arcangelis, L. & Reder, S. A random fuse model for breaking processes. *J. Phys. Lett.* **46**, L585–L590 (1985).
- Herrmann, H. J. & Roux, S. *Statistical Models for the Fracture of Disordered Media*. (North-Holland, Amsterdam, 1990).
- Herrmann, H. J., Hansen, A. & Roux, S. Fracture of disordered, elastic lattices in two dimension. *Phys. Rev. B* **39**, 637–648 (1989).
- Malthe-Sørensen, A. et al. Simulation of extensional clay fractures. *Phys. Rev. E* **58**, 5548–5564 (1998).
- Yakobson, B. I. Morphology and rate of fracture in chemical decomposition of solids. *Phys. Rev. Lett.* **67**, 1590–1593 (1991).
- Farver, J. R. & Yund, R. A. Oxygen bulk diffusion measurements and TEM characterization of a natural ultramylonite: implications for fluid transport in mica-bearing rocks. *J. Metamorph. Geol.* **17**, 669–683 (1999).
- Roark, R. J. & Young, W. C. *Formulas for Stress and Strain* (McGraw Hill, Tokyo, 1975).
- Allen, D. M. de G. *Relaxation Methods* (McGraw-Hill, New York, 1954).

Acknowledgements

This work was funded by the Norwegian Research Council through a grant to the Strategic University Program ‘Fluid Rock Interactions’. We thank J. Feder, P. Meakin and the other members of the Fluid Rock Interaction group for valuable suggestions and discussions and A. Kühn and M. Lund for providing some of the figures.

Correspondence and requests for materials should be addressed to B.J. (e-mail: bjorn.jamveit@geologi.uio.no).

Elevated CO₂ increases productivity and invasive species success in an arid ecosystem

Stanley D. Smith*, Travis E. Huxman*†, Stephen F. Zitzer‡, Therese N. Charlet*, David C. Housman*, James S. Coleman§, Lynn K. Fenstermaker||, Jeffrey R. Seemann‡ & Robert S. Nowak#

* Department of Biological Sciences, University of Nevada, Las Vegas, Nevada 89154-4004, USA

‡ Department of Biochemistry, University of Nevada, Reno, Nevada 89557, USA

§ Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, Nevada 89512, USA

|| Division of Earth and Ecosystem Sciences, Desert Research Institute, Las Vegas, Nevada 89119, USA

Department of Environmental and Resource Sciences, University of Nevada, Reno, Nevada 89557, USA

Arid ecosystems, which occupy about 20% of the earth's terrestrial surface area, have been predicted to be one of the most responsive ecosystem types to elevated atmospheric CO₂ and associated global climate change^{1–3}. Here we show, using free-air CO₂ enrichment (FACE) technology in an intact Mojave Desert ecosystem⁴, that new shoot production of a dominant perennial shrub is doubled by a 50% increase in atmospheric CO₂ concentration in a high rainfall year. However, elevated CO₂ does not enhance production in a drought year. We also found that above-ground production and seed rain of an invasive annual grass increases more at elevated CO₂ than in several species of native annuals. Consequently, elevated CO₂ might enhance the long-term success and dominance of exotic annual grasses in the region. This shift in species composition in favour of exotic annual grasses, driven by global change, has the potential to accelerate the fire cycle, reduce biodiversity and alter ecosystem function in the deserts of western North America.

The well documented increase in atmospheric CO₂ concentration ([CO₂]) acts to increase photosynthesis, plant biomass and plant water-use efficiency in many plant species⁵. Primary production in deserts is strongly limited by water and nitrogen resources³. Conceptual models predict that desert ecosystems will be the most responsive to increased [CO₂] because the strong response of water-use efficiency in plants alleviates water limitations to primary production^{1,2}. In addition, differences among species in their response to elevated [CO₂] can modify competitive interactions, potentially changing community composition. For example, the dominance of the exotic grass species *Bromus tectorum* (cheatgrass), which has invaded many thousand hectares in western North America, may be enhanced by elevated [CO₂] and thus alter competitive balance and the fire cycle in semi-arid shrub-steppe environments^{6–8}. If these predictions are realized, biodiversity in arid and semi-arid ecosystems could be significantly reduced, which in turn might alter ecosystem processes such as primary production, nutrient dynamics and landscape water balance⁹.

To examine the response of a desert ecosystem to elevated [CO₂], we established the Nevada Desert FACE (free-air CO₂ enrichment¹⁰) Facility (NDFF) in an undisturbed Mojave Desert ecosystem within the Nevada Test Site of southern Nevada, USA⁴. The climate is arid, with most precipitation occurring as rain during winter months. The NDFF is located within a desert scrub community (< 20% total perennial cover) dominated by the evergreen shrub *Larrea tridentata* (creosote bush) and several species of deciduous shrubs.

Perennial grasses are also common on the site, and dense stands of winter annuals occur in response to significant winter rains.

We have continuously maintained [CO₂] at a 550 µmol mol⁻¹ (p.p.m.) set point in three 23-m diameter experimental plots since April 1997, while maintaining three other plots at ambient [CO₂] and three control plots without the FACE apparatus. Elevated CO₂ treatment occurs 24 h d⁻¹, except during high winds (5-min average wind speed > 8 m s⁻¹) or at low air temperature (< 3 °C)⁴. CO₂ treatment occurred 93% of the time when air temperature exceeded 3 °C, and mean [CO₂] averaged 537 ± 41 and 375 ± 18 µmol mol⁻¹ in the elevated and ambient [CO₂] plots, respectively.

In the 1997–1998 hydrologic year (October to September) corresponding to the 1998 growing season (typically March to September for perennials; March to June for annuals), the southwestern United States experienced a pronounced El Niño cycle during which the NDFF received 309 mm of rain, which was 2.4-fold higher than the long-term average for the area. In contrast, the 1998–1999 hydrologic year was a dry La Niña cycle in which 107 mm of rain fell, and only 17 mm occurred before the onset of the growing season.

For the dominant perennial shrubs of the NDFF, we measured above-ground production during the 1998 and 1999 growing seasons. The cumulative increase in new-shoot biomass for *L. tridentata* was significantly higher (roughly double) at elevated [CO₂] than at ambient [CO₂] in 1998 ($P < 0.01$), but no difference in new shoot production occurred between CO₂ treatments in the dry year of 1999 ($P = 0.26$) (Fig. 1). In addition, no differences in shrub production occurred between the two types of control plots, and thus the FACE apparatus did not influence these production measurements. We also monitored new root production with minirhizotrons during the same growth period (data not shown) and saw no evidence that plants shifted to new root production in the dry year. Results from the deciduous shrubs at the NDFF, primarily *Ambrosia dumosa* and *Lycium andersonii*, also showed substantially enhanced new shoot production at elevated [CO₂] in 1998, but not in 1999 (data not shown).

In response to the anomalous 1997–1998 wet cycle, a large number of annual plants germinated at the NDFF during the winter months. Four species of annuals, *Bromus madritensis* ssp. *rubens*, *Eriogonum trichopes*, *Lepidium lasiocarpum* and *Vulpia*

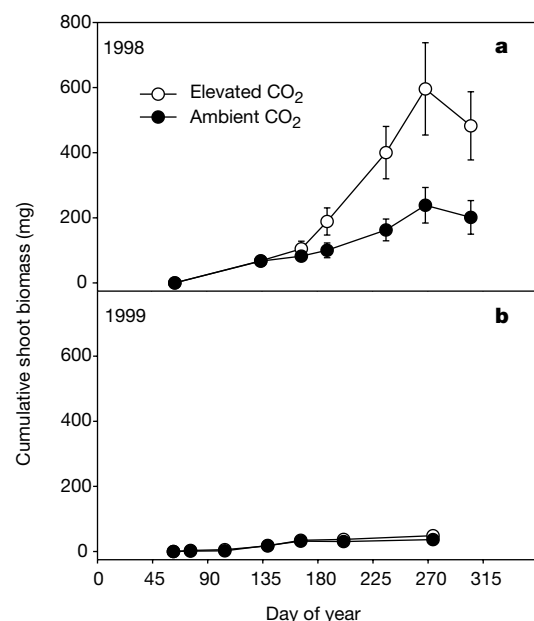


Figure 1 New shoot production of *Larrea tridentata* at ambient and elevated [CO₂] at the Nevada Desert FACE Facility in 1998 (a) and 1999 (b) ($n = 27$ for each CO₂ treatment).

† Present address: Department of Environmental, Population, and Organismal Biology, University of Colorado, Boulder, CO, 80309, USA.

octoflora, constituted over 70% of the total density of annuals in the plots and so were selected for above-ground harvests throughout the growth cycle. *Bromus* (red brome) is an exotic annual grass that has invaded many areas of the Mojave and Sonoran Deserts¹¹; the other three taxa are native annuals. Because of the pronounced spatial variation in soil nutrients that occurs in desert ecosystems¹², two primary microsites were chosen for harvest locations: (1) open interspaces between perennial plants; and (2) beneath or within the canopy of perennial plants. Results from the NDFP confirmed this 'fertile island' effect—soils beneath deciduous and evergreen shrubs have similar concentrations of inorganic nitrogen in the surface 10-cm layer (15 p.p.m. NO₃-N; 3 p.p.m. NH₄-N), and these concentrations are three- to fivefold higher than in the open interspaces (3 p.p.m. NO₃-N; 1 p.p.m. NH₄-N).

Total density of the native annuals (*Eriogonum*, *Lepidium* and *Vulpia*) at peak biomass (May) in 1998 was 42% lower ($P < 0.01$) and total above-ground biomass was 40% higher at elevated [CO₂] ($P < 0.1$) than at ambient [CO₂] (Fig. 2a). The net result was that individual plants were 2.4-times larger at elevated [CO₂] (0.94 g per plant) than at ambient [CO₂] (0.39 g per plant) ($P < 0.01$). For *Bromus*, both plant density and biomass increased at elevated [CO₂] (Fig. 2a). The 2.3-fold increase in *Bromus* above-ground biomass at elevated [CO₂] ($P < 0.05$) was due to a 50% increase in density coupled with a 53% increase in individual plant biomass. Early in the growth season (March), plant density did not differ among elevated [CO₂] and both types of control plots (data not shown), and thus the differences at peak biomass were not caused by differences in seed germination or in seed banks. Because none of the measured parameters differed significantly between the two types of control plots throughout the growing season, the FACE apparatus did not influence these results. Therefore, we conclude that as annual populations exhibited natural self-thinning from germination to peak biomass, both native annuals and *Bromus* responded to elevated [CO₂] with larger plants, but *Bromus* became a higher proportion of total plant density at elevated [CO₂]. In the dry winter–spring of 1999, there was no germination of annual plants and therefore no production in either the ambient or elevated [CO₂] plots.

Table 1 Seed number and mean seed mass for three annual species and two microsites at ambient and elevated [CO₂] in May, 1998 at the Nevada Desert FACE Facility

Taxa	Microsite	[CO ₂]*	Seed number (plant)	Seed number (LSA)	Seed mass
<i>Bromus</i>	Interspaces	Ambient	97	32	2.13
		Elevated	196*	42	1.95*
	Fertile-islands	Ambient	235	29	2.15
		Elevated	535*	32	1.91*
<i>Vulpia</i>	Interspaces	Ambient	107	108	0.43
		Elevated	138	86	0.51*
	Fertile-islands	Ambient	133	86	0.46
		Elevated	230*	64*	0.53
<i>Lepidium</i>	Interspaces	Ambient	200	90	0.33
		Elevated	634*	76	0.33
	Fertile-islands	Ambient	578	56	0.33
		Elevated	897*	79	0.34

Taxa are *Bromus madritensis* ssp. *rubens*, *Vulpia octoflora* and *Lepidium lasiocarpum*. Microsites are open interspaces between perennial plants and beneath perennial plant microsites. Seed number is expressed per plant and per unit leaf surface area (LSA). Seed mass is the mean mass for individual seeds (in mg).

* Parameters are significantly different from the data point immediately above (that is, response at elevated [CO₂] versus ambient [CO₂] for that species and cover type).

Bromus exhibited a threefold higher total seed rain at the plot level at elevated [CO₂] (Fig. 2a). Although seed rain across the whole plot by native annuals was not significantly increased by elevated [CO₂] (Fig. 2a), [CO₂] effects occurred at an individual plant level. For example, individuals of *Bromus* and *Lepidium* produced greater numbers of seeds at elevated than at ambient [CO₂], regardless of microsite location (Table 1). In contrast, *Vulpia* showed an enhancement in seed production only in fertile island microsites. As *Bromus* and *Lepidium* showed no changes in seed production per unit leaf surface area (LSA) and *Vulpia* showed a slight decrease per LSA, the increases in seed number occurred as a result of larger plants at elevated [CO₂]. However, individual seed mass decreased at elevated [CO₂] in *Bromus*, a phenomenon that did not occur for the native species (Table 1). *Bromus* seeds at elevated [CO₂] also have lower N content¹³, and the reduction in seed quality for *Bromus* at elevated [CO₂] results in a 20% reduction in growth rate of subsequent seedlings¹⁴. However, *Bromus* produces so many more seeds at

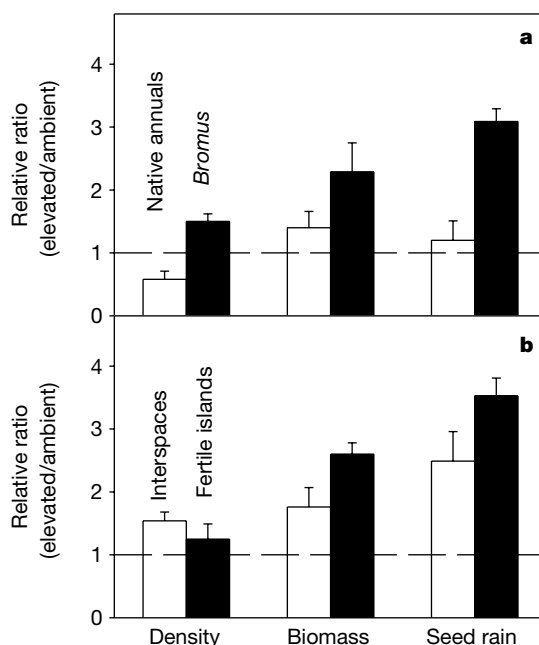


Figure 2 The relative ratio (elevated CO₂/ambient CO₂) of plant density (left), above-ground plant biomass (centre) and seed rain (total seed production; right) per unit area at the Nevada Desert FACE Facility in May, 1998. **a**, Native annuals (open bars) and *Bromus madritensis* ssp. *rubens* (solid bars). **b**, *Bromus madritensis* ssp. *rubens* in open

interspaces (open bars) and in 'fertile island' (solid bars) microsites. Note the dashed baseline at a ratio of 1; above the line plants at elevated CO₂ are greater than at ambient CO₂ for that parameter; below the line plants at ambient CO₂ are greater.

elevated $[CO_2]$ that it potentially compensates for reduced seed quality. This phenomenon might have important implications for population structure and competitive relationships between *Bromus* and native annuals, as well as affect consumers in an ecological system in which granivores have a fundamental role¹⁵.

Using *Bromus*, we found an interactive effect of elevated $[CO_2]$ and microsite on plant density, biomass and seed rain ($P < 0.01$) (Fig. 2b). Elevated $[CO_2]$ increased *Bromus* biomass in open interspaces by 76% but by twice that in fertile island microsites (160%). The stimulation of *Bromus* seed rain by elevated $[CO_2]$ in fertile islands (3.5-fold) was also greater than in interspaces (2.5-fold). In addition, we observed $[CO_2]$ -by-microsite interactions for the native annuals (data not shown). This interactive effect is consistent with predictions that desert annuals will show their strongest response to elevated $[CO_2]$ on sites with moderate to high soil fertility, whereas plant production on low-fertility desert soils is much less responsive to elevated $[CO_2]$ ^{1,3}.

These results have several important implications for the structure and function of desert ecosystems. First, total ecosystem above-ground primary production (both annuals and perennials) during an anomalously wet year for this arid region was substantially higher at elevated $[CO_2]$. Desert ecosystems have been predicted to exhibit a 50% increase in primary production with a doubling of $[CO_2]$, an increase that would be higher than for any other global biome type². Results from our FACE study exceeded this prediction in a wet year (a roughly 50% increase in native annual plant production, a doubling of new shoot production in the dominant shrub and a 2.3-fold increase in production of an exotic annual grass occurred with a 50% increase in $[CO_2]$). This also exceeds a roughly 25% increase in primary production observed in a young, rapidly growing pine forest in North Carolina using a similar FACE experimental design¹⁶. However, elevated $[CO_2]$ did not enhance productivity at the NDFF in a drought year. Apparently, plant physiological processes are sufficiently constrained by water stress during dry years in the desert such that plant production is minimal at both $[CO_2]$. Indeed, our plant gas-exchange results indicate increased photosynthesis and reduced stomatal conductance (and therefore transpiration) at elevated $[CO_2]$ in *Larrea* during wet seasons, but not during drought^{17,18}. By enhancing water-use efficiency in plants to the greatest extent when soil moisture is most abundant in this water-limited system, elevated $[CO_2]$ stimulates an already high rate of production in a wet year, but not in a dry year. Desert ecosystems have extremely high temporal variability in rainfall and primary production¹⁹, and our observations of a wet-dry cycle in the Mojave Desert suggest that elevated $[CO_2]$ will increase temporal variability of ecosystem production cycles.

Second, an invasive exotic annual grass showed significantly higher plant density, biomass and seed rain at elevated $[CO_2]$ than did several species of native annuals. The growth of a closely related exotic *Bromus* (*B. tectorum*) is also particularly responsive to elevated $[CO_2]$ ⁶. The greater response of exotic *Bromus* species to elevated $[CO_2]$ relative to native species could enhance the success of a group of highly invasive species in western North America. *B. tectorum* has invaded many thousands of hectares in this region²⁰, which has increased fire frequency²¹ and has converted large expanses of sagebrush-steppe vegetation to a fire-controlled annual grassland²². Post-fire competitive superiority of *Bromus* seedlings then precludes the reestablishment of perennial species²³. *Bromus madritensis* ssp. *rubens* has initiated a similar invasion cycle in the Mojave and Sonoran Deserts¹¹, although not yet to the extent as the invasion of *B. tectorum* farther north. The results from this study, showing both higher total biomass and seed rain in *Bromus* than in native annuals, confirm experimentally in an intact ecosystem that elevated $[CO_2]$ may enhance the invasive success of *Bromus* spp. in arid ecosystems. Enhanced success of *Bromus* spp. would expose these deserts to an accelerated fire cycle to which they are not adapted, and thus convert long-lived shrublands to annual grass-

lands dominated by exotic grasses throughout the region. □

Methods

Above-ground shrub production

We determined above-ground shoot production in the dominant perennial shrubs (*Larrea tridentata*, *Ambrosia dumosa*, *Lycium andersonii*) by monitoring incremental changes in stem length over each growing season. Three shoot tips were marked, proximal to the newest emerged leaf pair, before each growing season on three shrubs per species per treatment plot. Total shoot length was measured at least monthly from the mark to the shoot tip (including any branching) with calipers. The number of leaves distal to the mark was also counted. Tissue harvests were used to create treatment-specific regression equations for each species that converted shoot length to total shoot biomass (dry weight). A repeated measures two-way analysis of variance (ANOVA) was used to evaluate the effect of time and $[CO_2]$ for each year for perennial plant growth.

Annual plant density and production

We quantified annual plant density in each plot with two permanently placed 8×0.2 m transects positioned randomly from the centre of the plot to the periphery. Transects were divided into 0.5×0.2 m sub-plots in which all rooted annual plants were counted by species. We also categorized data by microsite type, allowing calculation of plant density in both open interspaces and within fertile island microsites (beneath evergreen shrubs, beneath deciduous shrubs and within perennial grass canopies). We determined biomass of the four dominant species (*Bromus madritensis* ssp. *rubens*, *Eriogonum trichopes*, *Lepidium lasiocarpum*, *Vulpia octoflora*) during five harvests at 2-week intervals from 7 April to 1 June 1998. At each harvest date, and for each microsite location within each plot, five plants of each species were harvested. We calculated plot-level annual biomass by multiplying plant density by mean individual plant biomass for each microsite, and then summing those values and correcting them for the proportion of microsite in each plot. For each response variable, a plot mean was used as the statistical unit. We determined the effect of elevated $[CO_2]$ on total annual plant density and biomass production with a one-way ANOVA. The effect of elevated $[CO_2]$ and microsite and their interaction for these parameters was determined individually for native and invasive annuals by a two-way ANOVA. Pairwise comparisons were determined by Student-Newman-Keuls tests ($\alpha = 0.05$).

Received 12 May; accepted 15 August 2000.

- Strain, B. R. & Bazzaz, F. A. in *CO₂ and Plants: The Response of Plants to Rising Levels of Carbon Dioxide* (ed. Lemon, E.) 177–222 (AAAS, Washington DC, 1983).
- Melillo, J. M. *et al.* Global climate change and terrestrial net primary production. *Nature* **363**, 234–240 (1993).
- Smith, S. D., Monson, R. K. & Anderson, J. E. *Physiological Ecology of North American Desert Plants* (Springer, Berlin, 1997).
- Jordan, D. N. *et al.* Biotic, abiotic and performance aspects of the Nevada Desert Free-Air CO₂ Enrichment (FACE) Facility. *Global Change Biol.* **5**, 659–668 (1999).
- Bazzaz, F. A. The response of natural ecosystems to rising global CO₂ levels. *Annu. Rev. Ecol. Syst.* **21**, 167–196 (1990).
- Smith, S. D., Strain, B. R. & Sharkey, T. D. Effects of CO₂ enrichment on four Great Basin grasses. *Funct. Ecol.* **1**, 139–143 (1987).
- Mayeux, H. S., Johnson, H. B. & Polley, H. W. in *Ecology and Management of Annual Grasslands* (eds Monsen, S. B. & Kitchen, S. G.) 95–100 (US Forest. Serv. Gen. Tech. Rep. INT-GTR-313, Ogden, Utah, USA, 1994).
- Sage, R. F. in *Carbon Dioxide, Populations and Communities* (eds Körner, C. & Bazzaz, F. A.) 231–249 (Academic, San Diego, 1996).
- Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
- Hendrey, G. R. & Kimball, B. A. The FACE program. *Agric. For. Meteorol.* **70**, 3–14 (1994).
- Hunter, R. *Bromus* invasions on the Nevada Test Site: present status of *B. rubens* and *B. tectorum* with notes on their relationship to disturbance and altitude. *Great Basin Nat.* **51**, 176–182 (1991).
- Schlesinger, W. H., Raikes, J. A., Hartley, A. E. & Cross, A. F. On the spatial pattern of soil nutrients in desert ecosystems. *Ecology* **77**, 364–374 (1996).
- Huxman, T. E., Hamerlynck, E. P. & Smith, S. D. Reproductive allocation and seed production in *Bromus madritensis* ssp. *rubens* at elevated CO₂. *Funct. Ecol.* **13**, 769–777 (2000).
- Huxman, T. E. *et al.* The effects of parental CO₂ on seed quality and subsequent seedling performance in *Bromus rubens*. *Oecologia* **114**, 202–208 (1998).
- Brown, J. H. & Heske, E. J. Control of a desert-grassland transition by a keystone rodent guild. *Science* **250**, 1705–1707 (1990).
- DeLucia, E. H. *et al.* Net primary production of a forest ecosystem with experimental CO₂ enrichment. *Science* **284**, 1177–1179 (1999).
- Hamerlynck, E. P. *et al.* Photosynthetic responses of *Larrea tridentata* to a step-increase in atmospheric CO₂ at the Nevada Desert FACE Facility. *J. Arid Environ.* **44**, 425–436 (2000).
- Huxman, T. E. *et al.* Photosynthetic down-regulation in *Larrea tridentata* exposed to elevated atmospheric CO₂: interaction with drought under glasshouse and field (FACE) exposure. *Plant Cell Environ.* **21**, 1153–1161 (1998).
- Noy-Meir, I. Desert ecosystems: environment and producers. *Annu. Rev. Ecol. Syst.* **4**, 51–58 (1973).
- Mack, R. N. The invasion of *Bromus tectorum* L. into western North America: an ecological chronicle. *Agro-Ecosystems* **7**, 145–165 (1981).
- Knapp, P. A. Spatio-temporal patterns of large grassland fires in the Intermountain West, U.S.A. *Global Ecol. Biogeogr. Lett.* **7**, 259–272 (1998).
- Young, J. A. & Evans, R. A. Population dynamics after wildfires in sagebrush grasslands. *J. Range Manage.* **31**, 283–289 (1978).
- Melgoza, G., Nowak, R. S. & Tausch, R. J. Soil water exploitation after fire: competition between *Bromus tectorum* (cheatgrass) and two native species. *Oecologia* **83**, 7–13 (1990).

Acknowledgements

We thank D. Jordan, E. Knight, G. Hendrey and the Brookhaven National Laboratory for technical support; C. Biggart, L. DeFalco, C. Grant, E. Hamerlynck, M. Hargrove, K. Hightower, K. Huxman, F. Landau, D. Pataki, K. Salsman, T. Shiver, D. Stortz-Lintz, and M. Taub for field and laboratory assistance; and J. Titus for soil nitrogen data. Financial support was provided by the DOE- and NSF-EPSCoR programs and panel grants from NSF/TECO and DOE/TCP, and additional support was provided by the NAES. We also thank the DOE/Nevada Operations Office and Bechtel Nevada for site operations support.

Correspondence and requests for materials should be addressed to S.D.S. (e-mail: ssmith@ccmail.nevada.edu).

Evolutionary origins of vertebrate appendicular muscle

C. Neyt[†], K. Jagla[‡], C. Thisse[§], B. Thisse[§], L. Haines^{*} & P. D. Currie^{*†}

^{*} Comparative and Developmental Genetics Section, Medical Research Council, Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

[‡] INSERM U.384, 28, Place Henri Dunant, 63001 Clermont Ferrand Cedex, France

[§] Institut de Génétique Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP 163, 67404 ILLKIRCH Cedex, France

[†] These authors contributed equally to this work

The evolution of terrestrial tetrapod species heralded a transition in locomotor strategies. While most fish species use the undulating contractions of the axial musculature to generate propulsive force, tetrapods also rely on the appendicular muscles of the limbs to generate movement^{1,2}. Despite the fossil record generating an understanding of the way in which the appendicular skeleton has evolved to provide the scaffold for tetrapod limb musculature³, there is, by contrast, almost no information as to how this musculature arose. Here we examine fin muscle formation within two extant classes of fish. We find that in the teleost, zebrafish, fin muscles arise from migratory mesenchymal precursor cells that possess molecular and morphogenetic identity with the limb muscle precursors of tetrapod species. Chondrichthyan dogfish embryos, however, use the primitive mechanism of direct epithelial somitic extensions to derive the muscles of the fin. We conclude that the genetic mechanism controlling formation of tetrapod limb muscles evolved before the Sarcopterygian radiation.

Within the zebrafish fin bud, the first indication of a cell's commitment to the muscle lineage begins with the expression of the myogenic transcription factor *myoD*. Expression of *myoD* initiates within the fin mesenchyme at 29 hours post-fertilization (h.p.f.), about 5 h after myotomal muscle differentiation is complete (Fig. 1a)^{4–6}. Expression is confined to the fin, with no detectable *myoD*-positive cells extending from adjacent somites (Fig. 1a). By 36 h.p.f., two distinct regions of *myoD* expression are evident, corresponding to the future dorsal and ventral muscle masses (Fig. 1b). Differentiation of fin muscle begins 12 h later, where antibodies against myosin heavy chain (MyHC) reveal unfused myocytes within muscle masses surrounding a condensing chondrocyte core (Fig. 1c). Muscle differentiation is discrete within the extending fin bud (Fig. 1c), and by 72 h elongated myofibrils can be seen appearing to attach to individual actinotrichia of the extending fin blade (Fig. 1d). The timing and morphogenesis of muscle differentiation in the zebrafish fin argues against the derivation of these cells from direct epithelial somite extensions—a mechanism generally invoked to describe the formation of the fin musculature in all fish species^{1,2,7–11}. We there-

fore endeavoured to detect the presence or absence of fin-level somitic extensions by direct microscopic examination. Time-lapse recording of somitic cells adjacent to the fin failed to detect the presence of epithelial extensions (Fig. 1e–h). Cells were observed instead to exit the somite at somitic boundaries, a few proliferative mesenchymal-like cells at a time, moving towards the developing fin bud (Fig. 1g, h).

To determine at what somitic level, if any, fin muscle precursors arise, the lipophilic cell tracer DiI was injected into individual somites of 24 h.p.f. (26-somite stage) embryos of a transgenic strain of zebrafish that expresses green fluorescent protein (GFP) under the control of the muscle-specific α -actin promoter. This promoter drives GFP expression in all differentiating skeletal muscle cells, including that of the fin¹². Coincident DiI (red) and GFP (green) fluorescence in the fin at 48 h.p.f. thus reveals a contribution of precursors derived from the initial DiI injection site to the fin musculature (Fig. 2a, b). This analysis indicated that cells derived from somites 2–4 may migrate to form the fin musculature (Fig. 2a, b; $n = 10$). Contribution to the fin musculature was only detected when DiI was injected adjacent to the most ventral cells of the

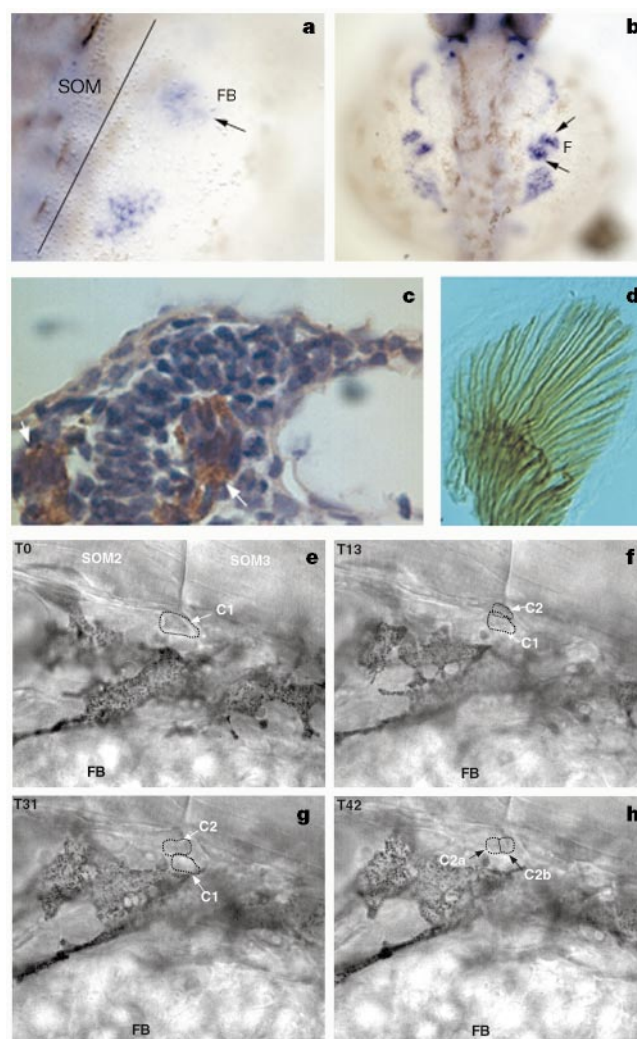


Figure 1 Zebrafish fin muscle formation. **a**, *myoD* expression initiates within the fin bud (arrow, FB) at 29 h.p.f. SOM, somites. **b**, *myoD* fin (F, arrows) expression at 36 h.p.f. **c**, Fin cross-section at 48 h.p.f. stained for MyHC (brown). Arrows indicate dorsal and ventral muscle masses. **d**, MyHC expression in the fin blade at 72 h.p.f. **e–h**, Cell movements from fin-level somites. Boundary between somites 2–3 of a 26-somite (24 h.p.f.) embryo in an oblique lateral view. T indicates time in minutes. C2a and C2b are daughters of cell C2.

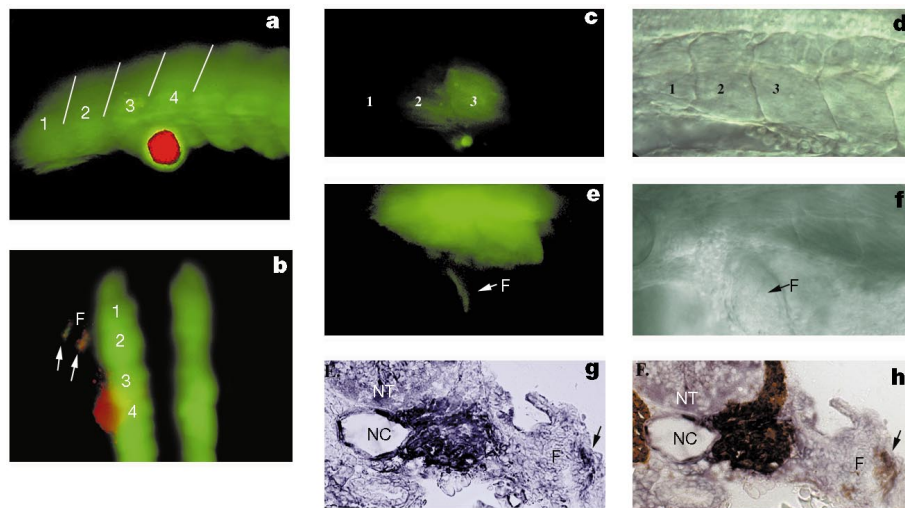


Figure 2 Identification of fin muscle precursors within zebrafish somites. **a**, Twenty-six-somite (24 h.p.f.), α -actin GFP transgenic embryo with ventral somite 4 labelled with Dil. **b**, Identical embryo at 48 h.p.f. revealing coincident Dil and GFP fluorescence within fin muscle masses. **c**, Uncaging of fluorescein dextran within somites 2 and 3 at 24 h.p.f.

d, DIC view of embryo in **c**. **e**, Identical embryo at 48 h.p.f. reveals labelled cells within the fin. **f**, DIC view of embryo in **e**. **g**, **h**, Co-immunolocalization of uncaged fluorescein dextran (**g**) and MyHC (**h**) in a fin level section of the identical embryo. F, fin; NC, notochord; NT, neural tube.

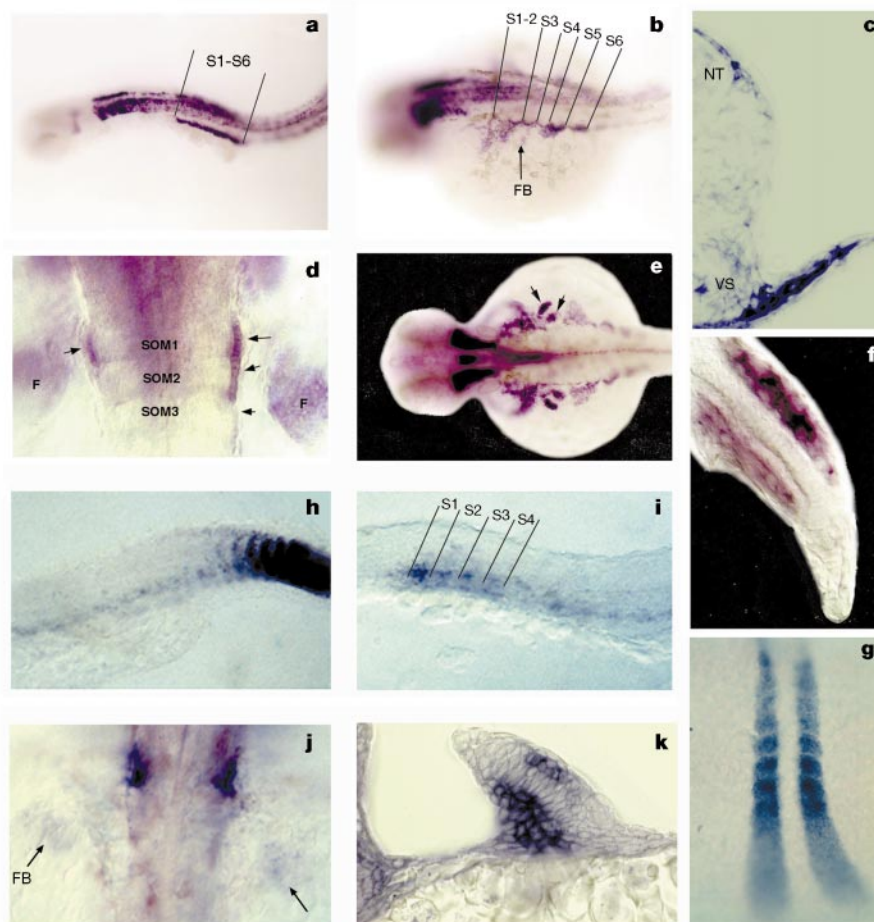


Figure 3 Expression of zebrafish *lbx1* and *max* genes. **a–f**, *lbx1* expression. **a**, At 22 somites, expression is restricted to ventral regions of somites (S) 1–6, the hindbrain and caudal regions of the neural tube. **b**, At 30 h.p.f., expression is also detected in cells emanating from individual somites and in the fin bud (FB, arrow). **c**, **d**, At 36 h.p.f., transcripts are restricted to the most ventral and lateral somitic (SOM) cells and to cells migrating from the somites (**c**, cross-section). NT, neural tube; VS, ventral somite. **e**, At

48 h.p.f., expression is evident in the differentiating fin muscle masses (F, arrows). **f**, At 72 h.p.f., expression is within fin muscle fibres. **g–k**, *max* expression. **g**, At 8 somites, expression is in the presomitic and somitic mesoderm. **h**, At 22 somites, transcripts are restricted to ventral regions of rostral somites. **i**, At 24 h.p.f., expression is upregulated in somites(s) 1–4. **j**, At 30 h.p.f., expression is in the fin bud (FB, arrows). **k**, At 48 h.p.f., expression is in differentiating fin muscle masses.

somite. DiI injections into dorsal somitic regions or other structures adjacent to the somite, such as the neural tube, never resulted in DiI-positive cells in the fin ($n = 12$).

Somitic cells were also labelled by the uncaging of caged fluorescein dextran within specific rostral somites¹³. Initially, groups of somites were uncaged in tandem (Fig. 2c–h), and when cells were identified that entered the fin the analysis was consecutively narrowed to individual somites and specific somite regions (see Supplementary Information). Contribution to the fin musculature was confirmed by co-immunolocalization of uncaged fluorescein and MyHC within the fin bud (Fig. 2g, h). The results of these experiments confirm that the ventral regions of somites 2–4 contribute cells to the fin musculature, with somite 4 contributing most fin muscle precursors (Fig. 2; see Supplementary Information).

Amniote limb myoblasts also derive from the migration of mesenchymal precursor cells from the ventro-lateral or hypaxial region of limb level somites¹⁴. We therefore wondered whether the genes required for correct formation of limb muscles in the mouse embryo would be similarly expressed in zebrafish fin myoblasts. Collectively, the homeobox-containing genes *lhx1* and *mox2* are expressed within the lateral hypaxial domain of mouse limb level somites, and this expression is maintained in actively migrating muscle precursor cells and continues in differentiating limb muscles after migration is complete^{15–19}. The introduction of null mutations in these genes results in a lack of specific muscle populations of the mouse limb^{16–19}. The identification of zebrafish homologues of the mouse *lhx1* and *mox2* genes reveal that they are expressed in specific regions of fin-level somites, in presumptive migratory myoblasts and in the developing musculature of the fin bud (Fig. 3; see

Supplementary Information). The expression patterns of these genes reinforce the notion that the muscles of the teleost pectoral fin arise from a small number of precursor cells present in the rostral somites that migrate to populate the fin in a manner directly analogous to limb muscle formation in amniotes.

The model of embryonic fin muscles deriving from direct extensions of epithelial muscle buds is based largely on studies carried out at the turn of the century on selacians^{1,7,8}. In these analyses, muscle buds were shown to extend from the ventral ends of myotomes adjacent to the forming fin, invade the developing fin mesenchyme and differentiate as elongated muscle fibres while remaining in contact with the myotome. Although evidence has cast doubt on the long-held belief that chondrichthyan species are primitive within the vertebrate phylogeny^{20–22}, both the articulation of the fin endoskeleton of selacians and its corresponding musculature are clearly primitive in comparison with the teleost fin³. Thus, we considered that the epithelial somitic extensions described to exist in this species might represent the primitive morphogenesis of embryonic fin muscles.

We therefore re-examined the formation of embryonic fin muscles in the selacian *Scyliorhinus canicula*. Beginning at the 15-mm stage, somites of pectoral, pelvic and interfin level possess an altered morphology in comparison with other somites (Fig. 4a, b; and data not shown). MyHC-positive cells can be detected outside the ventral somitic region, but adjacent to it (Fig. 4a). As development proceeds, differentiating myocytes can be detected extending from the somite in increasingly ventral positions (Fig. 4c, d). Eventually, lines of myocytes are detected in the fin bud mesenchyme and at interfin levels they extend ventrally to generate lines of differentiated muscles that populate the most ventral extent of the body wall

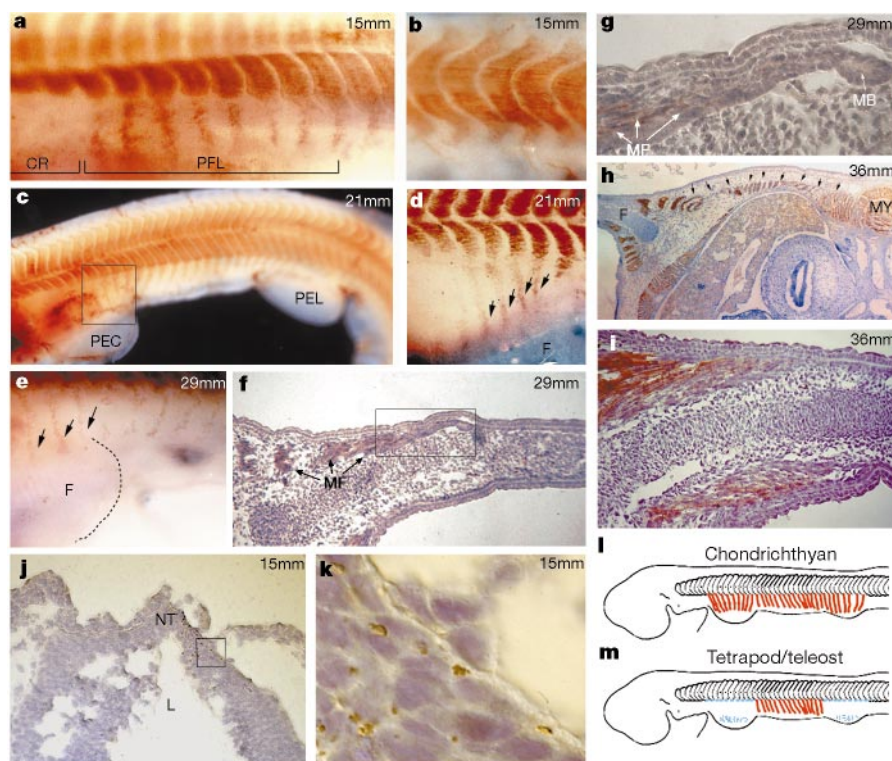


Figure 4 Fin muscle formation in *S. canicula*. **a, b**, 15-mm embryo stained for MyHC. PFL, pectoral fin level; CR, cranial somites; **b**, tail somites. **c**, 21-mm embryo stained for MyHC. PEC, pectoral fin; PEL, pelvic fin. **d**, Magnification of the region boxed in **c** (F, Fin). **e**, 29-mm embryo. MyHC-positive cells (arrows) extend to the pectoral fin. **a–e** are lateral views anterior to the left. **f**, 29-mm embryo. Fin cross-section stained for MyHC. MF, muscle fibres. **g**, Magnification of the area boxed in **f**, revealing the epithelial nature of the extending muscle buds (MB). **h**, 36-mm embryo. Cross-section stained for MyHC. Muscle

fibres (arrows) extend from the myotome (MY) to the fin (F). Dorsal to the right. **i**, 36-mm embryo. Fin-bud cross-section stained for MyHC. **j, k**, Hindbrain cross-section of a 15-mm embryo incubated with an anti-LBX antibody. L, lumen; NT, neural tube. **l**, Chondrichthyan embryos use extending epithelial muscle buds (red) to generate fin and body wall muscle. **m**, Tetrapod and teleost species share a common migratory mechanism for appendicular muscle formation (blue dots) and deploy the primitive mechanism of epithelial extensions to generate body wall muscle (red).

(Fig. 4e–g). The epithelial nature of the muscle bud can clearly be recognized in fin-level sections at this stage, with proximal cells of the extending bud differentiating first as MyHC-expressing myocytes while in contact with the distal cells of the epithelial muscle bud which differentiate later (Fig. 4f, g). At mature fin stages, contribution from several fin-level somites results in lines of muscle cells that extend from the myotome into the fin, dorsal and ventral to condensing cartilaginous extensions (Fig. 4h, i). Coincident with the differentiation of extended muscle fibres in the fin, epithelial muscle buds are no longer evident (Fig. 4i). This analysis reveals that the fin muscles of *S. canicula* derive from epithelial extensions of fin-level somites. Epithelial muscle buds have also been observed within the fins of the chimaera, *Callorhinchus milli* (C.N., D.A. Didier-Dagit and P.D.C., unpublished data). *C. milli* is believed to be the most basal extant member of the Holocephali²³, the most primitive of gnathostomes. The existence of epithelial muscle buds within this species therefore reinforces the generality of this primitive form of morphogenesis within the chondrichthyan clade.

We wondered whether a different genetic mechanism could underlie the altered morphogenesis of teleost and chondrichthyan fin musculature. To test this hypothesis, we cloned a fragment of the *S. canicula lbx* gene and performed whole-mount *in situ* hybridization on embryonic stages that span fin-muscle formation in this species. Although *lbx* expression could be detected in the hindbrain, neural tube and lateral mesoderm—known sites of *lbx* expression in zebrafish embryos—only general low-level somitic expression could be detected at early embryonic stages, before outgrowth of epithelial buds (Fig. 2; and data not shown). Furthermore, antibodies raised against LBX¹⁹, which crossreact with zebrafish LBX in a pattern identical to that of the transcript (data not shown), fail to detect LBX-positive myoblasts within *S. canicula* muscle buds at any stage examined (15 mm, 29 mm, 36 mm, $n = 5$), despite revealing LBX positive nuclei within the hindbrain (Fig. 4j, k). The lack of *lbx* expression in selacian muscle buds throughout their morphogenesis suggests that different molecular cues direct formation of fin muscles in teleost and chondrichthyan species.

In light of these results, it is interesting to contrast the formation of fin and body wall muscles in *S. canicula* to the formation of hypaxially derived intercostal and ventral body wall muscles in amniotes. Amniote interlimb somites possess an altered morphology to those at limb level, using structures termed ventral somitic buds to form the lateral dermomyotome. These buds remain epithelial, and give rise to the intercostal and body wall muscles by continuous ventral extension^{24–29}. The lateral somitic buds of interlimb somites do not express genes, such as *lbx1*, which have been implicated in the genetic control of limb myoblast migration. Thus, the interlimb-level somites of amniotes possess morphogenetic and molecular similarity with fin- and interfin-level somites of selacians (Fig. 4l, m). We postulate that the mechanisms used to generate hypaxial muscle from these somites are evolutionarily related, and represent the primitive condition for somite morphogenesis. This mechanism was genetically altered before the Sarcopterygian radiation, to create migratory myoblasts in fin-level somites that give rise to the appendicular musculature present in bony fishes and tetrapods. □

Methods

Immunohistochemistry and *in situ* hybridization

Whole-mount immunohistochemistry and *in situ* hybridization on zebrafish embryos were carried out as described^{6,30}. *S. canicula* embryos were removed from their egg casings and dissected from the yolk mass, and the length of embryos was recorded before fixation in 4% paraformaldehyde. Methods for *S. canicula* whole-mount immunohistochemistry were modified from zebrafish protocols⁶ incorporating an essential trypsinization step (0.5% v/w in PBS) up to 20 h at room temperature. Embryos were then 'cracked' in acetone at –20 °C for 15 min before incubation with the antibody A4-1025, and antibody binding was visualized by standard techniques⁶. We prepared cryostat sections as described⁶, and carried out antibody incubation and detection as described^{6,19}. Fin sections were taken from the pectoral level. We detected LBX protein on *S. canicula* sections by

serial sections through hindbrain to the trunk, inclusive of the pectoral fin, for 15-, 29- and 36-mm embryos. These included sections serial to those used for the detection of MyHC (29 and 36 mm). All sections were counterstained with eosin.

Time-lapse and cell-labelling strategies

Cell movements from somites were recorded in continuous time lapse to determine definitively the origin and relationship of cells. Cells were imaged on a Zeiss Axioskop under DIC optics, captured on a Hamamatsu ORCA-100 camera and analysed using IP lab software. Zebrafish embryos (24 h.p.f., 26 somites) derived from a transgenic strain in which GFP expression is driven by the zebrafish α -actin promoter¹² were microinjected at specific somitic levels with 0.1% (v/w) DiI (Sigma). Migration of cells from the labelled intrasomatic region was determined 24 h later (48–55 h.p.f.) and contribution to the fin musculature definitively determined by colocalization with DiI-positive cells with the GFP fluorescent cells of the fin musculature. Protocols for the use of caged fluorescein dextran (relative molecular mass 10,000; Molecular Probes) in fate mapping of fin level somites were derived from published methods¹³.

Identification of zebrafish and *S. canicula lbx* gene homologues

We identified a fragment of the zebrafish *lbx* gene, encoding the homeodomain, from complementary DNA pools prepared from 26-somite-stage embryos by using degenerate PCR. A similar fragment of the *S. canicula lbx* gene was isolated by degenerate PCR on cDNA pools created from 15-mm embryos. The *S. canicula lbx* fragment isolated encodes the identical amino acids to those encoded by the zebrafish gene and exhibits 86% sequence identity over a 166-bp region of the homeodomain. A full-length zebrafish *lbx* cDNA was isolated by hybridization of the zebrafish homeodomain-encoding fragment to a 26 somite gridded cDNA library (RZPD, Berlin). Alignment of zebrafish LBX to other identified vertebrate LBX proteins is included in the Supplementary Information.

Received 20 April; accepted 30 August 2000.

- Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* Vol. 1 (Dover Publications, New York/Constable and Company LTD, London, 1958).
- Kardong, K. V. *Vertebrates, Comparative Anatomy, Function, Evolution* (Wm C. Brown, Oxford, 1995).
- Coates, M. I. The origin of vertebrate limbs. *Development* (suppl.) *The Evolution of Developmental Mechanisms* 169–180 (1994).
- Weinberg, E. S. *et al.* Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development* **122**, 271–280 (1996).
- Devoto, S. H., Melancon, E., Eisen, J. S. & Westerfield, M. Identification of separate slow and fast muscle precursor cells *in vivo*, prior to somite formation. *Development* **122**, 3371–3380 (1996).
- Blagden, C. S., Currie, P. D., Ingham, P. W. & Hughes, S. M. Notochord induction of zebra fish slow muscle is mediated by Sonic Hedgehog. *Genes Dev.* **11**, 2163–2175 (1997).
- Braus, H. Beitrage zur entwicklung der muskulatur und des peripheren nervensystems der selachier. *Morphologisches Jahrbuch* **27**, 501–629 (1899).
- Dohrn, A. Die paarigen und unpaaren Flossen der Selachier. *Mitteilungen Zoologischen Station Neapel* **5**, 161–195 (1884).
- Corning, H. K. Über die ventralen urwirbelknospen in der brustflosse der teleostier. *Morphologisches Jahrbuch* **22**, 79–98 (1842).
- Harrison, R. G. Die entwicklung der unpaaren und paarigen flossen der teleostier. *Arch. Mikrosk. Anat. EntwMech.* **46**, 500–578 (1895).
- Grimm, M. Origin of the muscle blastemas in the developing pectoral fin of the rainbow trout (*Salmo gairdneri*). *Folia Morphologica* **21**, 197–199 (1973).
- Higashijima, S., Okamoto, H., Ueno, N., Hotta, Y. & Eguchi, G. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Dev. Biol.* **192**, 289–299 (1997).
- Kozłowski, D. J., Murakami, T., Ho, R. K. & Weinberg, E. S. Regional cell movement and tissue patterning in the zebrafish embryo revealed by fate mapping with caged fluorescein. *Biochem. Cell Biol.* **75**, 551–562 (1997).
- Christ, B., Jacob, H. J. & Jacob, M. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol.* **150**, 171–186 (1977).
- Jagla, K. *et al.* Mouse *Lbx1* and human *LBX1* define a novel mammalian homeobox gene family related to the *Drosophila lady bird* genes. *Mech. Dev.* **53**, 345–356 (1995).
- Mankoo, B. S. *et al.* *Mox2* is a component of the genetic hierarchy controlling limb muscle development. *Nature* **400**, 69–73 (1999).
- Schäfer, K. & Braun, T. Early specification of limb muscle precursor cells by the homeobox gene *Lbx1h*. *Nature Genet.* **23**, 213–216 (1999).
- Brohmann, H., Jagla, K. & Birchmeier, C. The role of *Lbx1* in migration of muscle precursor cells. *Development* **127**, 437–445 (2000).
- Gross, M. K. *et al.* *Lbx1* is required for muscle precursor migration along a lateral pathway into the limb. *Development* **127**, 413–424 (2000).
- Rasmussen, A. S. & Arnason, U. J. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. *Mol. Evol.* **48**, 118–123 (1999).
- Janvier, P. Major events in vertebrate evolution. *Trends Ecol. Evol.* **14**, 298–299 (1999).
- Basden, A. M., Young, G. C., Coates, M. I. & Ritchie, A. The most primitive osteichthyan braincase? *Nature* **403**, 185–188 (2000).
- Didier, D. A. Phylogenetic systematics of extant chimeroid fishes (Holocephali, Chimaeroidei). *Am. Mus. Novitates* **3119**, 1–86 (1995).
- Christ, B., Jacob, M. & Jacob, H. J. On the origin and development of the ventro-lateral trunk musculature in the avian embryo. An experimental and ultrastructural study. *Anat. Embryol. (Berl.)* **166**, 87–101 (1983).
- Christ, B. & Ordahl, C. P. Early stages of chick somite development. *Anat. Embryol. (Berl.)* **191**, 381–396 (1995).
- Cinnamon, Y., Kahane, N. & Kalcheim, C. Characterisation of the early development of specific hypaxial muscles from the ventrolateral myotome. *Development* **126**, 4305–4315 (1999).

27. Denetclaw W. F. & Ordahl, C. P. The growth of the dermomyotome and formation of early myotome lineages in thoracolumbar somites of chicken embryos. *Development* **127**, 893–905 (2000).
28. Dietrich, S., Schubert, F. R., Healy C., Sharpe, P. T. & Lumsden, A. Specification of the hypaxial musculature. *Development* **125**, 2235–2249 (1998).
29. Houzelstein, D. et al. The homeobox gene *Mx1* is expressed in a subset of somites, and in muscle progenitor cells migrating into the forelimb. *Development* **126**, 2689–2701 (1999).
30. Currie, P. D. & Ingham, P. W. Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature* **382**, 452–455 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We are indebted to B. Mankoo for his gift of the zebrafish *mox* clone and for sharing results before publication. We also thank S. Hughes and M. Goulding for antibodies; and M. Buckingham, R. Sporle, N. Rosenthal, N. Hastie, R. Currie and members of the Currie lab for discussion and critical reading of the manuscript. We thank the staff at the University of Glasgow's Milport Marine Station, Isle of Cumbrae, for the supply of *S. canicula* embryos; T. Chapman and J. Mattocks for animal husbandry; and P. Perry for help with microscopy. K.J. was supported by grants from the Human Frontier Science Program and from the Association Francaise contre les Myopathies (AFM). C.T. and B.T. were supported by funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Hôpital Universitaire de Strasbourg, the Association de Recherche sur le Cancer and the Ligue Nationale Contre le Cancer. P.D.C. is supported by a MRC career development fellowship and the BBSRC.

Correspondence and requests for materials should be addressed to P.D.C. (e-mail: petec@hgu.mrc.ac.uk).

Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA

Amy E. Pasquinelli^{*,†}, Brenda J. Reinhart^{*,†}, Frank Slack[‡], Mark Q. Martindale[§], Mitzi I. Kuroda^{||}, Betsy Maller[‡], David C. Hayward[¶], Eldon E. Ball[‡], Bernard Degnan[#], Peter Müller^{††}, Jürg Spring^{††}, Ashok Srinivasan^{**}, Mark Fishman^{**}, John Finnerty^{††}, Joseph Corbo^{‡‡}, Michael Levine^{‡‡}, Patrick Leahy^{§§}, Eric Davidson^{§§} & Gary Ruvkun^{*}

^{*} Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA

[‡] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA

[§] Kewalo Marine Lab, Pacific Biomedical Research Center, University of Hawaii, Honolulu, Hawaii 96813, USA

^{||} Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

[¶] Research School of Biological Sciences, PO Box 475,

Australian National University, Canberra, ACT 2601, Australia

[#] Department of Zoology and Entomology, University of Queensland, Brisbane, Queensland 4072 Australia

^{††} Institute of Zoology, University of Basel, CH-4051 Basel, Switzerland

^{**} Cardiovascular Research Center, Massachusetts General Hospital,

Charlestown, Massachusetts 02129, USA

^{‡‡} Department of Biology, Boston University, Boston, Massachusetts 02215, USA

^{‡‡} Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

^{§§} Division of Biology 156-29, California Institute of Technology, Pasadena, California 91125, USA

[†] These authors contributed equally to this work

Two small RNAs regulate the timing of *Caenorhabditis elegans* development^{1,2}. Transition from the first to the second larval stage fates requires the 22-nucleotide *lin-4* RNA^{1,3,4}, and transition from late larval to adult cell fates requires the 21-nucleotide *let-7* RNA². The *lin-4* and *let-7* RNA genes are not homologous to each

other, but are each complementary to sequences in the 3' untranslated regions of a set of protein-coding target genes that are normally negatively regulated by the RNAs^{1,2,5,6}. Here we have detected *let-7* RNAs of ~21 nucleotides in samples from a wide range of animal species, including vertebrate, ascidian, hemichordate, mollusc, annelid and arthropod, but not in RNAs from several cnidarian and poriferan species, *Saccharomyces cerevisiae*, *Escherichia coli* or *Arabidopsis*. We did not detect *lin-4* RNA in these species. We found that *let-7* temporal regulation is also conserved: *let-7* RNA expression is first detected at late larval stages in *C. elegans* and *Drosophila*, at 48 hours after fertilization in zebrafish, and in adult stages of annelids and molluscs. The *let-7* regulatory RNA may control late temporal transitions during development across animal phylogeny.

The sequence and function of both the *lin-4* and *let-7* small RNAs are conserved in the nematode *Caenorhabditis briggsae*^{1,2}. BLASTN searches reveal one DNA segment from the *Drosophila melanogaster* genome sequence, three segments from the human genome sequence on chromosomes 9, 11 and 22 bearing exact sequence matches, and two other human segments on chromosomes 9 and 21 with 20/21 matches to the *let-7* RNA (Fig. 1a). Database searches did not detect potential *lin-4* homologues in any genus except the *Caenorhabditidae*. Similar stem-loop secondary structures are predicted for precursor transcripts of *Caenorhabditidae*, *Drosophila* and human *let-7* RNAs (Fig. 1a). The mature 21-nucleotide (nt) *let-7* RNA may be efficiently processed from this putative precursor because only the mature RNA is detected in *C. elegans* and most

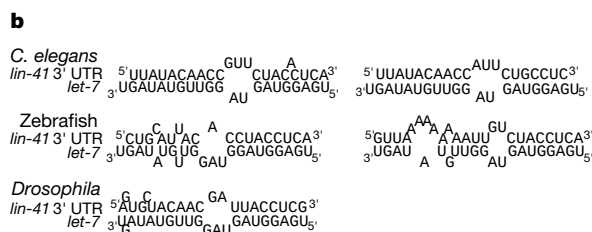
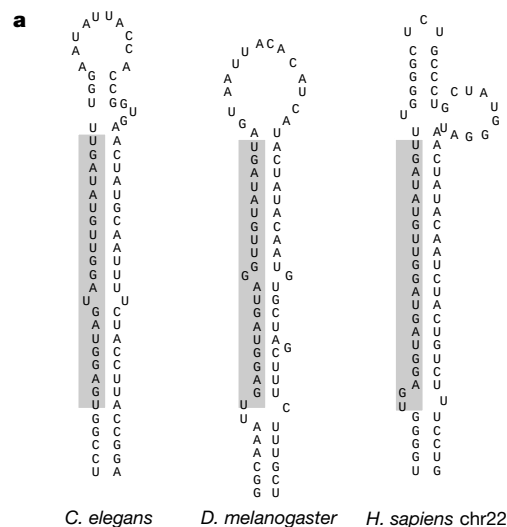


Figure 1 *let-7* gene sequences. **a**, Stem-loop structures of *C. elegans*, *D. melanogaster* and *Homo sapiens* *let-7* theoretical longer transcripts. The 21-nt *let-7* region is shaded. *let-7* genomic regions from *C. elegans* (Z70203), *D. melanogaster* (AE003659) and *H. sapiens* chromosome 22 (AL049853). The two human *let-7* homologous genes on chromosome 9 are tandemly arranged and separated by 369 base pairs; clustered with the human chromosome 22 *let-7* exact match is an 18/21 match to the *let-7* RNA. **b**, The 3' UTRs of the *D. melanogaster* (AA3990768) and *Danio rerio* *lin-41* (AI794385) cDNAs contain *let-7* complementary sites.

other species (see below). A similarly structured larger transcript is also predicted for the *lin-4* RNA, and rare transcripts that may correspond to it have been detected in *C. elegans*¹.

The *let-7* RNA regulates late developmental events in *C. elegans* by downregulating *lin-41* and perhaps other genes that contain sequences complementary to the small RNA in their 3' untranslated regions (UTRs)^{2,6}. *lin-41* encodes an RBCC protein that has *Drosophila* and vertebrate orthologues⁶. *let-7* complementary sites are present in the 3' UTRs of both the *Drosophila* and zebrafish *lin-41* complementary DNAs (Fig. 1b).

let-7 RNA transcripts of ~21 nt are expressed in human tissues and *Drosophila*. (Fig. 2). As in *C. elegans*, only RNAs in the ~21-nt range were detected. The expression levels of the human *let-7* RNA varied among tissues, indicating possible cell-type regulation of *let-7* expression. As *let-7* is expressed late in animal development, it may be significant that the lowest level of human *let-7* is observed in bone

marrow, which consists of a large proportion of immature cells. The *let-7* RNA from each human tissue could be the product of any or all of the several human *let-7*-like genes.

The *C. elegans* and human *let-7* RNAs were not detected with a sense strand probe and are RNase A sensitive (data not shown). Consistent with the RNA predicted by genome analysis, the sequences of the human and fly *let-7* RNAs appear identical to the *C. elegans let-7* RNA as assayed by S1 analyses (Fig. 2c). However, a single mismatch between the RNA and oligonucleotide could be tolerated in both S1 and northern analyses. Also consistent with the database searches, we did not detect *lin-4* RNA in non-nematode species (Figs 2 and 3).

The expression and function of the *let-7* RNA in *C. elegans* begins during the third larval stage, when the gene specifies a transition from late larval to adult cell fates, and continues at all subsequent stages². Expression of *Drosophila let-7* is also temporally regulated: *let-7* RNA is absent until the late third instar, just before metamorphosis (Fig. 2b). At the early pupal stage, there is at least a 10-fold increase in *let-7* RNA expression that is sustained to adulthood. Thus, temporal regulation of *let-7* is similar in *Drosophila* and *C. elegans*, which are distantly related members of the ecdysozoan clade⁷.

The size and temporal regulation of *let-7* are also conserved in the more distantly related lophotrochozoan clade, members of which do not moult but have larval and adult stages. For example, in two mollusc species and in a polychaete annelid, *let-7* RNA is expressed

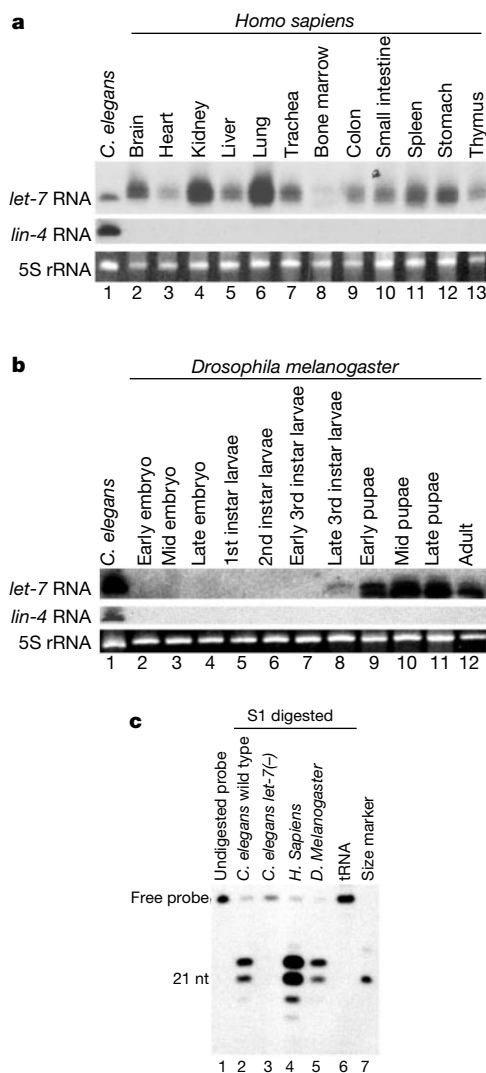


Figure 2 Expression of *let-7* RNA in human and *Drosophila*. **a**, Northern blot of total RNA from mixed stage *C. elegans* (lane 1) and human tissues (lanes 2–13) probed for *let-7* RNA and then stripped and re-probed for *lin-4* RNA. 5S rRNA serves as a loading control. **b**, Northern blot of total RNA from mixed stage *C. elegans* (lane 1) and *D. melanogaster* developmental stages (lanes 2–12), probed as in (a). **c**, S1 nuclease mapping detects similar transcripts in *C. elegans*, *Drosophila* and human total RNA. A 5'-end-labelled antisense strand probe undigested (lane 1) and digested after hybridization to RNA from wild-type *C. elegans* (lane 2), *C. elegans let-7(mn112)* null mutant (lane 3), human adult lung tissue (lane 4), *Drosophila* late pupal stage (lane 5) or transfer RNA (lane 6).

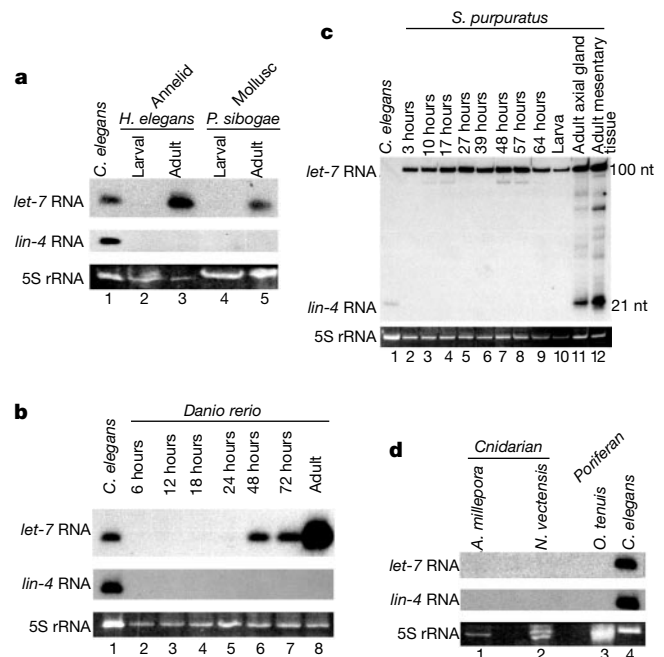


Figure 3 Expression of *let-7* RNA is developmentally regulated in lophotrochozoans and deuterostomes. **a**, Northern blot of total RNA from mixed stage *C. elegans* (lane 1), *Hydroides elegans* trocophore larvae (lane 2) and adults (lane 3), *Phestilla sibogae* veliger larvae (lane 4) and adults (lane 5), probed as in Fig. 2a. **b**, Northern blot of total RNA from mixed stage *C. elegans* (lane 1) and *D. rerio* developmental stages (lanes 2–8), probed as in Fig. 2a. **c**, Northern blot of total RNA from mixed stage *C. elegans* (lane 1) and *Strongylocentrotus purpuratus* embryonic stages 3–64 hours after fertilization (lanes 2–9), pluteus larvae (lane 10) and adult tissues (lanes 11–12), probed as in Fig. 2a. **d**, Northern blot of total RNA from adult *Acropora millepora* (lane 1), adult *Nematostella vectensis* (lane 2), adult *Ophiastorgia tenuis* (lane 3) and mixed stage *C. elegans* (lane 4), probed as in Fig. 2a. We detected a very weak 21-nt signal in a planula stage *Acropora millepora* but it was probably caused by contamination from a bilaterian predator because these animals are collected from open sea cultures and we could not repeat this detection in an independent planula sample or in embryonic stage *A. millepora*.

at the adult stage but not at larval stages (Fig. 3a). Thus, *let-7* may function at later stages of these species to regulate developmental progression to the adult.

Vertebrates do not develop through larval stages. But in zebrafish expression of *let-7* RNA is also temporally regulated: expression commences in the embryo between 24 and 48 hours after fertilization and continues with strong expression at the adult stage (Fig. 3b). Analyses of other deuterostomes shows expression of a 21-nt *let-7* RNA in other vertebrates, as well as in urochordate ascidians, and in a hemichordate (Fig. 4). In the echinoderm *Strongylocentrotus purpuratus*, *let-7* may be regulated at the level of precursor RNA processing rather than transcription; a ~100-nt *let-7* RNA is detected during embryonic and early larval development, but at the adult stage the 21-nt *let-7* RNA and possible processing intermediates appear (Fig. 3c).

We did not detect convincing *let-7* RNA in anthozoan and hydrozoan cnidarian species nor in two poriferan species (Figs 3d and 4). Plant and unicellular organisms also failed to show *let-7* RNA expression, consistent with database searches for those species that have been completely sequenced.

Because all three main clades of bilaterian animals express a *let-7* RNA that is temporally regulated, but cnidarian, poriferan and all non-animal species that we analysed do not express a detectable *let-7* RNA, we propose that the gene evolved after the divergence of diploblastic and bilaterian animals (Fig. 4). However, the lack of detection of *let-7* RNA in the simpler animal and non-animal species might also reflect sequence divergence or loss of the gene in the particular species tested, rather than lack of these genes in a common ancestor.

Although the observation that *let-7* homologous RNAs across phylogeny are temporally regulated does not prove that these RNAs

actually specify temporal patterning in this broad range of species, the conservation of sequence, of a longer structured precursor, of the 21-nt length, of temporal regulation and of complementary sites in the *lin-41* target in two ecdysozoan and one chordate species, are strong evidence of a conserved function. There are candidate mutations that map to the location of the *Drosophila let-7* and *lin-41* genes which may provide a way to test the function of these genes in another ecdysozoan.

The *let-7* RNA is conserved across bilaterian phylogeny but the earlier acting *lin-4* RNA is not. Consistent with this, the *let-7* target *lin-41* is conserved in *Drosophila* and vertebrates⁶, whereas the *lin-4* target *lin-14* appears to be unique to nematodes⁵. The *let-7*-regulated late larval transition in *C. elegans* may be ancestrally related to late transitions in other species, for example from larval to reproductive forms, whereas the earlier *lin-4/lin-14*-regulated transition may be a recent invention of the nematode phylum. Alternatively, *lin-4* and *lin-14* may evolve more quickly.

The 21-nt length of the *let-7* RNA is highly conserved, indicating that this size is central to its function. It may be significant that this length is similar to the 21–25-nt RNAs observed during RNA interference (RNAi)-directed downregulation of target messenger RNAs^{8,9}. However, there are differences between the mechanisms: *let-7* is encoded and only the sense strand is expressed, whereas the sense and antisense small RNAs involved in RNAi are processed from exogenously supplied double-stranded RNAs that are hundreds of base pairs long^{8–12}. RNAi degrades target mRNA^{8–12}, whereas the heterochronic *lin-4* RNA affects the translation but not the stability of its target mRNAs^{5,13}. Also, mutations in *C. elegans* genes that confer resistance to RNAi¹⁴ do not have heterochronic phenotypes indicative of a defect in *lin-4* or *let-7* regulation of target genes (C. Mello, personal communication). One common feature of RNAi and the putative precursors of the heterochronic RNAs is longer regions of duplex RNA (Fig. 1a), suggesting that similar processing and amplification pathways could generate these 21-nt RNAs.

Two ~21-nt RNAs regulate *C. elegans* temporal development, and we argue that one of these RNAs is likely to regulate developmental timing in bilaterian animals. We propose that these types of RNAs be called small temporal RNAs (stRNAs). Genome sequence comparisons and expression analyses among bilaterian animals may reveal additional stRNAs that regulate other developmental transitions. □

Methods

Human total RNA samples were purchased from Clontech. Total RNA preparation, northern analysis and S1 nuclease protection assays were performed as described². Oligonucleotides used as northern probes were *let-7* 5'-AACTATACAACTACTACCT-CACCGGATCC-3' and *lin-4* 5'-ATAGTACACTCACACTTGAGGTCTCAGGG-3'. 5S rRNA was detected by ethidium bromide staining of the polyacrylamide gels before transfer. Oligonucleotides used for S1 nuclease analyses were 5'-CTATACAACTAC-TACCTCACCGGAT-3' (3' end match to *C. elegans* genomic region), 5'-ACTATA-CACCTACTACTACCTCACCCCA-3' (3' end match to human genomic region) and 5'-ACTATACAACTACTACTCAATTGTC-3' (3' match to *Drosophila* genomic region).

Received 12 June; accepted 11 August 2000.

- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Reinhart, B. et al. The 21-nucleotide *let-7* RNA regulates developmental timing in *C. elegans*. *Nature* **403**, 901–906 (2000).
- Chalfie, M., Horvitz, H. R. & Sulston, J. E. Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell* **24**, 59–69 (1981).
- Ambros, V. & Horvitz, H. R. The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev.* **1**, 398–414 (1987).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
- Slack, F. et al. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
- Aguinaldo, A. M. et al. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
- Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
- Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).

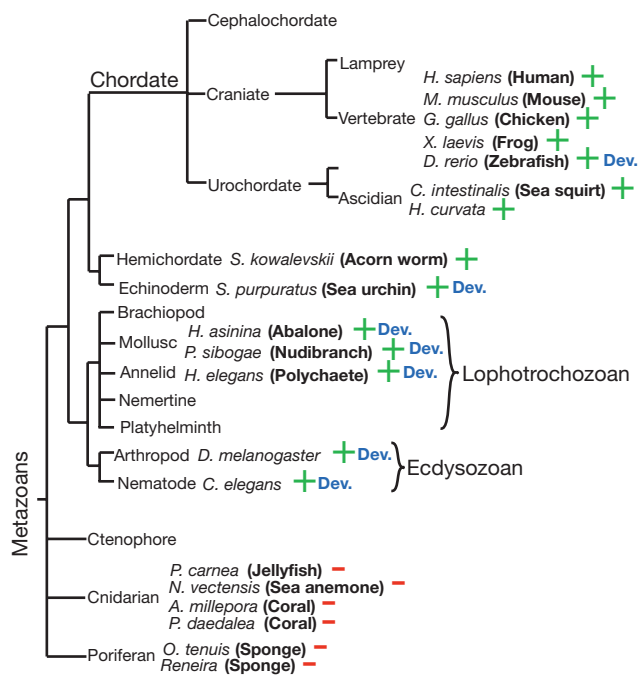


Figure 4 Phylogenetic comparison of *let-7* RNA expression. Phylogenetic tree showing species that do (+) and do not (–) express *let-7* RNA. Species in which we detect the conserved developmental pattern of *let-7* RNA expression (no *let-7* RNA in early stages but *let-7* expression by adulthood) are indicated by 'Dev.'. The full genus and species names are as follows: *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus laevis*, *Danio rerio*, *Ciona intestinalis*, *Herdmania curvata*, *Saccoglossus kowalevskii*, *Strongylocentrotus purpuratus*, *Haliotis asinina*, *Phestilla sibogae*, *Hydroids elegans*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Podocoryne carnea*, *Nematostella vectensis*, *Acropora millepora*, *Platygyra daedalea*, *Ophlitaspongia tenuis* and *Reneira* sp.

10. Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. Targeted mRNA degradation by double-stranded RNA *in vitro*. *Genes Dev.* **13**, 3191–3197 (1999).
11. Montgomery, M. K., Xu, S. & Fire, A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 15502–15507 (1998).
12. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–810 (1998).
13. Olsen, P. H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
14. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).

Acknowledgements

We thank the following people for RNA and tissue samples: T. Heanue, R. Pearce and C. Tabin for chick; J. Gerhart and M. Kirschner for *Xenopus* and acorn worm; S. Agarwal for *Xenopus*; N. Stavropoulos for mouse; C. Unabia and K. del Carmen for annelid and mollusc; H. Bode for Hydra; J. Nardone and S. Ferrari for *Arabidopsis*; P. Sudarsanam for yeast; and D. Selinger for *E. coli*. The phylogenetic survey in this work was inspired by the NASA Evolution and Development meetings organized by E. Davidson and C. Golden. This work was supported by an NIH grant from NIGMS to G.R. and a grant from the MGH Fund for Medical Discovery to A.E.P.

Correspondence and requests for materials should be addressed to G.R.

Polarization of the anterior–posterior axis of *C. elegans* is a microtubule-directed process

Matthew R. Wallenfang & Geraldine Seydoux

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

In *Caenorhabditis elegans*, polarity along the anterior–posterior (A/P) axis is established shortly after fertilization and is determined by the sperm, whose position specifies the posterior end of the embryo¹. Although many factors required for the establishment of A/P polarity have been described^{2,3}, the nature of the spatial cue provided by the sperm remains unknown. Here we show that a microtubule-organizing centre is necessary and sufficient to establish several aspects of A/P polarity. In wild-type embryos, appearance of the first molecular asymmetries along the A/P axis correlates with and requires nucleation of microtubules by the sperm-derived centrosomes (sperm asters). In mutant embryos arrested in meiosis, sperm asters fail to form, and posterior is defined by the position of the persistent meiotic spindle rather than by the position of the sperm. Together, our data indicate that the primary spatial cue for A/P polarity in *C. elegans* derives from microtubules emanating from the sperm asters. Our findings support a parallel^{4–7} between *C. elegans* zygotes and other cells, such as *Drosophila* oocytes, which rely on microtubules to regulate polarity.

Caenorhabditis elegans oocytes appear to have no predetermined A/P polarity¹. The first manifestations of polarity are seen roughly 30 min following fertilization, after the oocyte chromatin, which had been arrested in prophase of meiosis I, completes meiosis. At that time, several proteins become localized asymmetrically along the long axis of the embryo: in particular, PAR-2 localizes to the cortex nearest the sperm pronucleus (future posterior end), and PAR-3 localizes in a reciprocal pattern on the cortex opposite the sperm pronucleus² (future anterior end; Fig. 1a). Each of these is required to localize the PAR-1 kinase to the posterior cortex. PAR-1, in turn, is required for the asymmetric segregation of several factors in the cytoplasm, including PIE-1 and the germline-specific P granules which localize to the posterior² (Fig. 1a). Segregation of

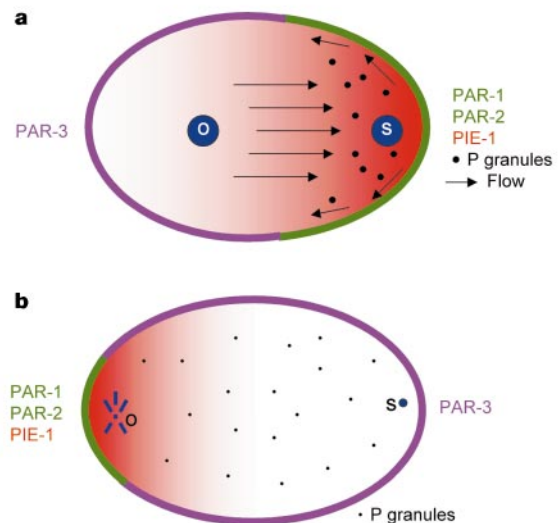


Figure 1 Establishment of A/P polarity in *C. elegans*. **a**, Polarity in wild-type embryos. Embryo is shown in first mitotic prophase during pronuclear migration. In most fertilization events, sperm entry occurs on the side opposite the oocyte nucleus¹. In this and Figs 2 and 3, the position of the oocyte chromatin is indicated by an 'O' and that of the sperm chromatin by an 'S' (in some figures, the sperm chromatin is not visible in the focal plane shown). Arrows indicate cytoplasmic flow, which flows towards the sperm pronucleus in the interior of the embryo and away from it along the cortex. Shown are the asymmetric distributions of PAR-1 and PAR-2 (green), PAR-3 (magenta), PIE-1 (red) and P granules (dots). **b**, Polarity in *mat* mutant embryos. The oocyte chromatin is arrested in metaphase of meiosis I and the sperm chromatin remains condensed.

P granules has been correlated with an internal flow of cytoplasm directed towards the sperm pronucleus, which also occurs during this period^{8,9} (Fig. 1a). The sperm pronucleus itself is dispensable for polarity¹⁰, leaving open the question of which sperm-derived component is responsible for polarizing the embryo.

To address this question, we have analysed a collection of temperature sensitive mutants (*mat-1*, *mat-2*, *mat-3*, *emb-27*, *emb-30*) that arrest shortly after fertilization while the oocyte chromatin is in metaphase of meiosis I (A. Golden, P. Sadler, M.R.W., G.S. and D. Shakes, manuscript in preparation). Consistent with their metaphase arrest phenotype, *mat-1* and *emb-30* encode *C. elegans* homologues of the anaphase-promoting complex (APC) components Cdc27 (J. Schumacher, D. Shakes and A. Golden, personal communication) and APC4/Lid1 (ref. 11), respectively.

Because fertilization usually occurs opposite the oocyte nucleus, mutant embryos in this collection arrest with the oocyte chromatin in meiosis at one end of the embryo and condensed sperm chromatin at the other end (Fig. 1b). To determine whether A/P polarity can be established under these conditions, we determined the distribution of PIE-1, which in wild type segregates towards the sperm pronucleus after meiosis is completed. We observed asymmetric PIE-1 in 48% of *mat-1* mutant embryos (Table 1). Notably, when asymmetric, PIE-1 always was enriched on the side of the oocyte chromatin rather than on the sperm side (Fig. 2a). Similar results were obtained with the other mutants (Table 1). Examination of live embryos expressing a PIE-1–GFP (green fluorescent protein) fusion showed that PIE-1 becomes asymmetric in all *mat-1* embryos (Fig. 2b); however, this asymmetry is transient, causing it to be observed in only 48% of embryos in asynchronous populations (Table 1).

Immunostaining with PAR-1, PAR-2 and PAR-3 antibodies showed that PAR asymmetry is also 'reversed' in *mat-1* mutants: PAR-1 and PAR-2, which in wild type localize to the cortex nearest the sperm, were on the cortex nearest the oocyte chromatin in *mat-1*

embryos (Table 1; and data not shown); and PAR-3, which in wild type localizes to the cortex opposite the sperm, was on the cortex opposite the oocyte chromatin in *mat-1* embryos (Table 1). As in wild type^{12,13}, PIE-1 asymmetry was dependent on *par-1* and *par-3* activities in *mat-1* mutants, and PAR-3 asymmetry was independent of *par-1* (Table 1). These observations indicate that, even though A/P polarity appears 'reversed' in *mat-1* embryos, the PAR hierarchy required to establish polarity in wild type is maintained in these mutants.

In contrast to the PARs and PIE-1, P granules remained uniformly distributed in *mat-1* embryos (Table 1). P granules in these mutants remained small and dispersed (Fig. 2c), as they appear in wild-type oocytes and newly fertilized embryos. Cytoplasmic flow, which in wild type has been correlated with P-granule segregation⁹, was also not observed in *mat-1* embryos (4 out of 4 embryos; see Methods). We conclude that, although completion of meiosis is not required for the establishment of PAR and PIE-1 asymmetries, it is required for P-granule segregation and cytoplasmic flow.

To understand the basis of the apparent reversal of PAR/PIE-1 asymmetry in *mat-1* embryos, we examined the distributions of PAR-2 and PAR-3 with respect to the location of the meiotic spindle. We found that localization of these proteins always was correlated (PAR-2) or inversely correlated (PAR-3) with the position of the meiotic spindle, even when the meiotic spindle was located off centre relative to the pole (Fig. 2d–f). In rare embryos (<1%), an additional patch of PAR-2 was sometimes observed away from the meiotic spindle. In all cases (5 out of 5), these ectopic patches were associated with foci of microtubules that appeared to have formed spontaneously in the cytoplasm, away from the oocyte and sperm chromatin (Fig. 2g).

These observations indicated that the meiotic spindle, rather than the sperm, may act as the main polarizing cue in embryos arrested in meiosis. To test this hypothesis further, we examined the consequence of disrupting the meiotic spindle in *mat-1* embryos. Treatment of *mat-1* embryos with the microtubule-depolymerizing drug nocodazole disrupted the integrity of the meiotic spindle (see Methods) and reduced the number of embryos with asymmetric PAR-2 (Table 1). To block formation of the meiotic spindle completely, we used RNA-mediated interference (RNAi) against *ncc-1*, a *C. elegans* cdc2 homologue required for entry into M phase¹⁴. *mat-1;ncc-1* embryos did not form a meiotic spindle and did not segregate PIE-1 (Table 1). These data are consistent with the meiotic spindle acting as the primary polarizing cue in *mat-1* mutants.

This finding raised the possibility that an eccentrically located

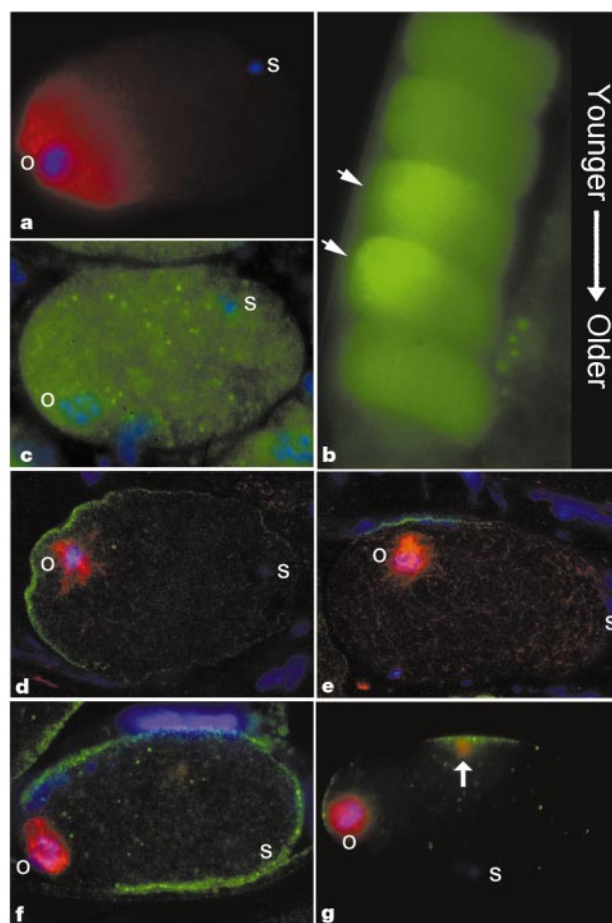


Figure 2 A/P polarity in *mat* mutant embryos. **a**, *mat-1(ax227)* embryo immunostained for PIE-1 (red) and DAPI (blue). **b**, Embryos inside the uterus of a *mat-1(ax227)* hermaphrodite expressing PIE-1–GFP (green). Embryos are arranged from youngest to oldest (top to bottom). All are in the one-cell stage; those with asymmetric PIE-1 are indicated by arrows. A similar pattern was observed in all hermaphrodites examined ($n > 50$), indicating that PIE-1 segregates transiently in all mutant embryos. **c**, *mat-1(ax227)* embryo immunostained for P granules (green) and DAPI (blue). This same embryo segregated PIE-1 towards the side of the oocyte chromatin (not shown). **d, e**, *mat-1(RNAi)* embryos expressing PAR-2–GFP (green) immunostained for α -tubulin (red) and DAPI (blue). **f**, *mat-1(ax227)* embryo immunostained for PAR-3 (green), α -tubulin (red) and DAPI (blue). **g**, *mat-1(ax227)* embryo immunostained for PAR-2 (green), α -tubulin (red) and DAPI (blue). Arrow points to an ectopic focus of microtubules.

Table 1 Asymmetric segregation in *mat* mutants

Protein scored	Genotype/treatment	% Asymmetric* (n)
PIE-1	<i>mat-1(ax227)</i>	48 (114)
PIE-1	<i>mat-2(ax76)</i>	39 (52)
PIE-1	<i>mat-3(ax148)</i>	45 (64)
PIE-1	<i>emb-27(ax81)</i>	42 (99)
PIE-1	<i>emb-30(ax69)</i>	38 (60)
PIE-1	<i>mat-1(RNAi)</i>	46 (84)
PIE-1	<i>mat-1(RNAi); par-1(b274)</i>	0 (27)
PIE-1	<i>mat-1(RNAi); par-3(e2074)</i>	3 (136)
PAR-2	<i>mat-1(ax227)†</i>	39 (247)
PAR-2	<i>mat-1(ax227); nocodazole</i>	21 (113)
PAR-3	<i>mat-1(RNAi)</i>	48 (42)
PAR-3	<i>mat-1(RNAi); par-1(b274)</i>	43 (21)
P granules‡	<i>mat-1(ax227)</i>	0 (32)
PIE-1:GFP	<i>mat-1(ax227)</i>	45 (55)
PIE-1:GFP	<i>mat-1(ax227); ncc-1(RNAi)</i>	0 (66)
PIE-1:GFP	<i>mat-1(ax227); air-1(RNAi)</i>	46 (67)

* Per cent of embryos in which the indicated protein/marker was asymmetric. All other embryos showed uniform distribution. When asymmetric, the protein assayed always was found enriched on the side of the oocyte chromatin (as identified by DAPI staining), except for PAR-3 which segregated to the side opposite the oocyte chromatin.

† These embryos were prepared in the same manner as the nocodazole treated embryos (see Methods) except that nocodazole was omitted.

‡ P granules could not be detected in 55% of *mat-1(ax227)* embryos. Only embryos with visible P granules were included in our analysis.

microtubule-organizing centre might be sufficient to polarize the *C. elegans* embryo along its A/P axis. In wild-type embryos, the sperm-derived centrosomes begin to nucleate microtubules around the sperm pronucleus (sperm asters) after the completion of meiosis¹⁵, and thus potentially could function as the polarity cue^{1,8,10}. To explore this hypothesis, we stained wild-type embryos expressing a PAR-2–GFP fusion with an antibody against α -tubulin (Fig. 3a–d). During meiosis (Fig. 3a), and immediately afterwards before sperm aster formation (Fig. 3b), PAR-2 was present at low levels uniformly throughout the cytoplasm and weakly at the cortex. Asymmetric PAR-2 was first detected on the cortex nearest the male pronucleus when sperm asters first became visible (Fig. 3c). A similar correlation was observed with PAR-1, PAR-3 and PIE-1 (data not shown). During sperm aster growth, the domain occupied by PAR-2 on the cortex expanded until reaching close to 50% embryo length by pronuclear meeting (data not shown). In later stages, during formation of the mitotic spindle, PAR-2 did not extend further (Fig. 3d), perhaps limited by the presence of PAR-3 in the anterior^{12,16}.

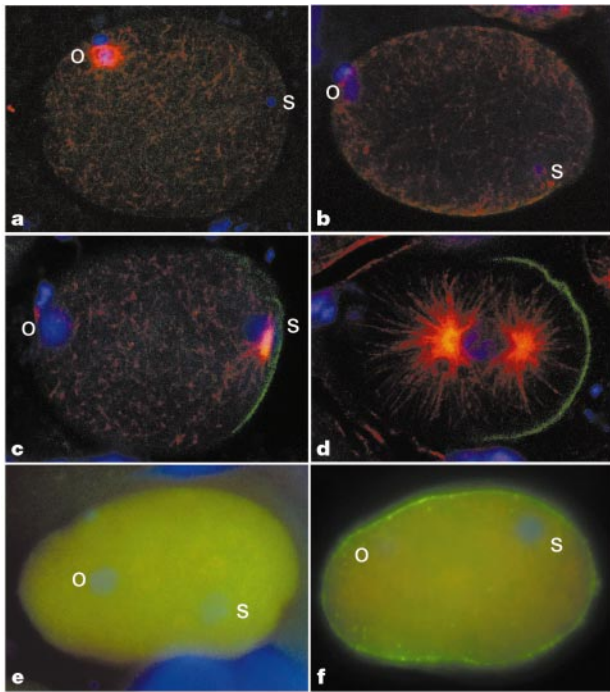


Figure 3 Establishment of A/P polarity in wild type correlates with and requires sperm aster formation. **a–d**, Wild-type one-cell embryos expressing PAR-2–GFP (green) immunostained for α -tubulin (red) and DAPI (blue). **a**, Meiosis II; **b**, end of meiosis; **c**, sperm aster formation; **d**, mitotic pro-metaphase (oocyte and sperm pronuclei have fused). **e**, *air-1(RNAi)* embryo expressing PIE-1–GFP (green) and immunostained for α -tubulin (red, diffuse due to lack of sperm asters) and DAPI (blue). **f**, *air-1(RNAi)* embryo immunostained for α -tubulin (red, diffuse), PAR-3 (green) and DAPI (blue).

Previous studies using microtubule inhibitors failed to determine a role for sperm asters in the establishment of polarity^{8,17}. Those studies, however, were performed in one-cell embryos that had completed pronuclear formation, and thus potentially did not address the possibility that sperm asters might be involved in establishing polarity at an earlier stage^{8,17}. To test this possibility, we disrupted sperm aster formation by interfering with *air-1* expression. *air-1* belongs, along with *Drosophila* Aurora and *Saccharomyces cerevisiae* Ipl1, to a subfamily of serine/threonine kinases that regulate spindle dynamics during mitosis¹⁸. Disruption of *air-1* expression by RNAi has been reported to disrupt PIE-1 localization¹⁸. We found that this defect correlates with a block in the formation of sperm asters. *air-1(RNAi)* embryos could be divided into two classes: embryos with no detectable asters (42%, 33/79); and embryos with small but distinct asters (58%, 46/79). Among the first class, 91% of embryos failed to localize PIE-1 (Fig. 3e). These embryos also had uniform PAR-3 at the cortex confirming that A/P polarity had not been established (Fig. 3f). In contrast, 80% of *air-1(RNAi)* embryos that had retained asters segregated PIE-1 and PAR-3. To confirm that the *air-1* A/P polarity defect was due to a lack of sperm asters, we examined the requirement for *air-1* in *mat-1* embryos where polarity is established independently of sperm asters. *mat-1;air-1* embryos were able to segregate PIE-1 to the same extent as *mat-1* embryos (Table 1), confirming that AIR-1 itself is not required to establish A/P polarity. We conclude that establishment of A/P polarity in wild-type embryos is dependent on sperm aster formation. A similar conclusion has been reached following a study¹⁹ showing that mutations in *spd-2*, which delay and attenuate sperm aster formation, eliminate A/P polarity in wild type.

Together, our results show that a microtubule-organizing centre is necessary, and can be sufficient, to establish many aspects of A/P

polarity in *C. elegans* embryos. We propose that when microtubules emanating from such a centre reach the cortex, they, or factors associated with them, affect the local balance between PAR-2 and PAR-3, perhaps by promoting binding of PAR-2 to the cortex and/or by inhibiting binding of PAR-3. This interaction may be dependent on the actin cytoskeleton, which is required for A/P polarization^{17,20}. We propose that under normal circumstances only microtubules nucleated by the sperm asters reach the cortex. Meiotic microtubules remain confined to the spindle, unless allowed to persist for a prolonged period as occurs in embryos arrested in meiosis. Indeed, meiotic spindles in *mat* mutants have a characteristic frayed appearance not seen in wild type (compare Fig. 2d–g with Fig. 3a).

One question raised by our studies is whether the initial point of contact between microtubules and cortex serves as a nucleation site from which posterior markers can spread, or whether microtubule–cortex interactions are required all along the posterior half of the embryo to define the complete PAR-2 domain. In support of the latter possibility, we have noted that PAR-2 domains are smaller in *mat* mutants than in wild type (compare Fig. 2d, e with Fig. 3d), suggesting that sperm aster growth is required for expansion of the PAR-2 domain. Alternatively or additionally, PAR-2 domain expansion could depend on actin-based mechanisms that fail to occur in *mat* embryos, such as cytoplasmic flow. Indeed, small PAR-2 domains have also been observed in embryos deficient for the non-muscle myosin II regulatory light chain, *mhc-4* (ref. 21), suggesting that refinement of PAR domains may require mechanisms distinct from those used for their establishment.

A homologue of PAR-1 has been shown to function in the establishment of A/P polarity in *Drosophila* oocytes^{4,5}, a process dependent on polarization of the microtubule cytoskeleton²². PAR-1-like kinases also exist in mammals where they have been implicated in regulating microtubule dynamics and epithelial cell polarity^{23,24}. Together with these studies, our finding that A/P polarity in *C. elegans* is initiated by microtubules supports the idea that conserved microtubule and PAR-1 dependent mechanisms regulate polarity in cells as diverse as *C. elegans* zygotes, *Drosophila* oocytes and mammalian epithelial cells. □

Methods

Strains and alleles

We used the following strains: N2 (wild type), *axIs73*(pJH3.92, pRF4), *axEx1094*(pMW1.03, pRF4), *mat-1(ax227ts)I*, *mat-1(ax227ts)II*; *axIs73*(pJH3.92, pRF4), *mat-2(ax76ts)II*, *mat-3(ax148ts)III*, *emb-27(ax81ts)II*, *emb-30(ax69ts)III*, *rol-4(sc8) par-1(b274)IV/nT1* (IV; V), *lon-1(e185) par-3(e2074) III*; and *sDp3* (III; f). *axIs73* is an integrated array containing a PIE-1–GFP transgene²⁵. *axEx1094* is an extrachromosomal array containing a PAR-2–GFP fusion under the control of the PIE-1 promoter. This transgene is expressed in early embryos in the same pattern as endogenous PAR-2 (data not shown). We isolated the *ax* alleles used in this study in a large-scale screen for temperature-sensitive embryonic lethal mutations (M.R.W. and G. S., unpublished data). Embryos arrested in meiosis were derived from mutant hermaphrodites that had been shifted as adults to the non-permissive temperature (25 °C) for 8–12 h. Under these conditions, only oocyte meiosis is affected; sperm meiosis, which in hermaphrodites occurs during larval development, is not affected.

Immunofluorescence microscopy

We used the following antibodies: anti-PIE-1 (ref. 13), anti-PAR-1 (ref. 26), anti-PAR-2 (ref. 16), anti-PAR-3 (ref. 12), anti-P granule²⁷ and anti- α -tubulin (mAbE7, Developmental Studies Hybridoma Bank; or DM1A, Sigma). Immunostaining was done as described¹³ using DAPI (4',6-diamidino-2-phenylindole dihydrochloride) to visualize oocyte and sperm DNA. We carried out analysis using either a compound fluorescence microscope (Zeiss Axioplan 2) or laser scanning confocal microscope (Zeiss LSM 510).

RNA interference

Double-stranded RNAi was used to knock out gene function using the bacterial feeding method²⁸. Targeted DNAs were cloned into the feeding vector L4440, transformed into *Escherichia coli* H115(DE3), and the resulting transformants were used to spread NGM plates containing 60 μ g ml⁻¹ ampicillin and 80 μ g ml⁻¹ IPTG. Bacterial lawns grown overnight were used immediately or stored at 4 °C for up to 3 weeks. Hermaphrodites were placed on bacterial lawns as L4 larvae and allowed to feed for 20–30 h at 25 °C before their embryos were examined.

Drug treatments

Gravid adult hermaphrodites were placed onto slides into 8 μ l of egg salts²⁹ containing 10 μ g per ml nocodazole (Sigma) in DMSO (0.1% final concentration) and cut open with a syringe needle to release embryos. Slides were incubated in a humid chamber for 15–20 min before processing for immunostaining with anti-PAR-2 and anti- α -tubulin antibodies as described above. This treatment severely disrupted the meiotic spindle in most embryos but was not sufficient to eliminate it entirely (data not shown).

Time-lapse videomicroscopy

Young adult hermaphrodites were anaesthetized in 0.1% tricaine, 0.01% tetramisole in M9 (ref. 30), placed on a 2% agarose pad and covered with a cover slip. We recorded time-lapse movies of embryos *in utero* from fertilization to first cleavage (~45 min) for wild-type embryos, and from fertilization to 75 min after fertilization for *mat-1* embryos, using a Zeiss Axioplan 2 equipped with a DAGE video camera and Scion LG3 frame grabber. Nomarski images were acquired every 3 s using the 4D grabber software (Laboratory for Optical and Computational Instrumentation, University of Wisconsin, Madison). Resulting Quicktime movies were analysed for directed cytoplasmic flow by following the movement of yolk granules as described⁸. In two of the 4 *mat-1* embryos analysed, we also examined PAR-2–GFP fluorescence at the end of the movie and confirmed that it had become asymmetric.

Received 10 May; accepted 16 August 2000.

- Goldstein, B. & Hird, S. N. Specification of the anteroposterior axis in *Caenorhabditis elegans*. *Development* **122**, 1467–1474 (1996).
- Rose, L. S. & Kemphues, K. J. Early patterning of the *C. elegans* embryo. *Annu. Rev. Genet.* **32**, 521–545 (1998).
- Bowerman, B. Maternal control of pattern formation in early *Caenorhabditis elegans* embryos. *Curr. Top. Dev. Biol.* **39**, 73–117 (1998).
- Shulman, J. M., Benton, R. & St Johnston, D. The *Drosophila* homolog of *C. elegans* PAR-1 organizes the oocyte cytoskeleton and directs oskar mRNA localization to the posterior pole. *Cell* **101**, 377–388 (2000).
- Tomancak, P. *et al.* A *Drosophila melanogaster* homologue of *Caenorhabditis elegans* par-1 acts at an early step in embryonic-axis formation. *Nature Cell Biol.* **2**, 458–460 (2000).
- Kemphues, K. PARing embryonic polarity. *Cell* **101**, 345–348 (2000).
- Morris, J., Lehmann, R. & Navarro, C. PARallels in axis formation. *Science* **288**, 1759–1760 (2000).
- Hird, S. N. & White, J. G. Cortical and cytoplasmic flow polarity in early embryonic cells of *Caenorhabditis elegans*. *J. Cell Biol.* **121**, 1343–1355 (1993).
- Hird, S. N., Paulsen, J. E. & Strome, S. Segregation of germ granules in living *Caenorhabditis elegans* embryos: cell-type-specific mechanisms for cytoplasmic localisation. *Development* **122**, 1303–1312 (1996).
- Sadler, P. L. & Shakes, D. C. Anucleate *Caenorhabditis elegans* sperm can crawl, fertilize oocytes and direct anterior-posterior polarization of the 1-cell embryo. *Development* **127**, 355–366 (2000).
- Furuta, T. *et al.* EMB-30: An APC4 homologue required for metaphase-to-anaphase transitions during meiosis and mitosis in *Caenorhabditis elegans*. *Mol. Biol. Cell* **11**, 1401–1419 (2000).
- Etemad-Moghadam, B., Guo, S. & Kemphues, K. J. Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. *Cell* **83**, 743–752 (1995).
- Tenenhaus, C., Schubert, C. & Seydoux, G. Genetic requirements for PIE-1 localization and inhibition of gene expression in the embryonic germ lineage of *Caenorhabditis elegans*. *Dev. Biol.* **200**, 212–224 (1998).
- Boxem, M., Srinivasan, D. G. & van den Heuvel, S. The *Caenorhabditis elegans* gene *ncc-1* encodes a cdc2-related kinase required for M phase in meiotic and mitotic cell divisions, but not for S phase. *Development* **126**, 2227–2239 (1999).
- Albertson, D. G. Formation of the first cleavage spindle in nematode embryos. *Dev. Biol.* **101**, 61–72 (1984).
- Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T. & Kemphues, K. J. PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. *Development* **122**, 3075–3084 (1996).
- Strome, S. & Wood, W. B. Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* **35**, 15–25 (1983).
- Schumacher, J. M., Ashcroft, N., Donovan, P. J. & Golden, A. A highly conserved centrosomal kinase, AIR-1, is required for accurate cell cycle progression and segregation of developmental factors in *Caenorhabditis elegans* embryos. *Development* **125**, 4391–4402 (1998).
- O'Connell, K., Maxwell, K. & White, J. The *spd-2* gene is required for polarization of the anteroposterior axis and formation of the sperm asters in the *Caenorhabditis elegans* zygote. *Dev. Biol.* **221**, 55–70 (2000).
- Hill, D. P. & Strome, S. Brief cytochalasin-induced disruption of microfilaments during a critical interval in one-cell *C. elegans* embryos alters the partitioning of developmental instructions to the two-cell embryo. *Development* **108**, 159–172 (1990).
- Shelton, C. A., Carter, J. C., Ellis, G. C. & Bowerman, B. The nonmuscle myosin regulatory light chain gene *mlc-4* is required for cytokinesis, anterior-posterior polarity, and body morphology during *Caenorhabditis elegans* embryogenesis. *J. Cell Biol.* **146**, 439–451 (1999).
- van Eeden, F. & St Johnston, D. The polarisation of the anterior-posterior and dorsal-ventral axes during *Drosophila* oogenesis. *Curr. Opin. Genet. Dev.* **9**, 396–404 (1999).
- Bohm, H., Brinkmann, V., Drab, M., Hense, A. & Kurzhals, T. V. Mammalian homologues of *C. elegans* PAR-1 are asymmetrically localized in epithelial cells and may influence their polarity. *Curr. Biol.* **7**, 603–606 (1997).
- Drewes, G., Ebner, A., Preuss, U., Mandelkow, E. M. & Mandelkow, E. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. *Cell* **89**, 297–308 (1997).
- Reese, K. J., Dunn, M. A., Waddle, J. A. & Seydoux, G. Asymmetric segregation of PIE-1 in *C. elegans* is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. *Mol. Cell* **6**, 445–455 (2000).

- Guo, S. & Kemphues, K. J. par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
- Strome, S. & Wood, W. B. Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **79**, 1558–1562 (1982).
- Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
- Edgar, L. G. Blastomere culture and analysis. *Methods Cell Biol.* **48**, 303–321 (1995).
- McCarter, J., Bartlett, B., Dang, T. & Schedl, T. On the control of oocyte meiotic maturation and ovulation in *Caenorhabditis elegans*. *Dev. Biol.* **205**, 111–128 (1999).

Acknowledgements

We are grateful to A. Golden, P. Sadler, J. Schumacher and D. Shakes for their characterization of the *mat* mutants. We also thank K. O'Connell and J. White for sharing results before publication; K. Kemphues and L. Boyd for antibodies; L. Timmons and A. Fire for RNAi feeding materials; and A. Golden, P. Sadler, D. Shakes, K. Kemphues, K. O'Connell, Y. Zheng and members of the Seydoux lab for comments on the manuscript. Some strains used in this study were provided by the *Caenorhabditis* Genetics Center (University of Minnesota). M.R.W. was an NSF predoctoral fellow. This work was supported by grants from the Searle and Packard Foundations to G.S.

Correspondence and requests for materials should be addressed to G.S. (e-mail: gseydoux@jhmi.edu).

Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors

Jun Yamashita*, Hiroshi Itoh*, Masanori Hirashima†, Minetaro Ogawa†, Satomi Nishikawa†, Takami Yurugi*, Makoto Naito‡, Kazuwa Nakao* & Shin-ichi Nishikawa†

* Department of Medicine and Clinical Science, and † Department of Molecular Genetics, Kyoto University Graduate School of Medicine, 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8509, Japan
‡ The Second Department of Pathology, Niigata University School of Medicine, Asahimachi-dori 1, Niigata 951-8510, Japan

Interaction between endothelial cells and mural cells (pericytes and vascular smooth muscle) is essential for vascular development and maintenance^{1–4}. Endothelial cells arise from Flk1-expressing (Flk1⁺) mesoderm cells⁵, whereas mural cells are believed to derive from mesoderm, neural crest or epicardial cells and migrate to form the vessel wall^{6–8}. Difficulty in preparing pure populations of these lineages has hampered dissection of the mechanisms underlying vascular formation. Here we show that Flk1⁺ cells derived from embryonic stem cells can differentiate into both endothelial and mural cells and can reproduce the vascular organization process. Vascular endothelial growth factor promotes endothelial cell differentiation, whereas mural cells are induced by platelet-derived growth factor-BB. Vascular cells derived from Flk1⁺ cells can organize into vessel-like structures consisting of endothelial tubes supported by mural cells in three-dimensional culture. Injection of Flk1⁺ cells into chick embryos showed that they can incorporate as endothelial and mural cells and contribute to the developing vasculature *in vivo*. Our findings indicate that Flk1⁺ cells can act as 'vascular progenitor cells' to form mature vessels and thus offer potential for tissue engineering of the vascular system.

Flk1, one of the receptors for vascular endothelial growth factor (VEGF), is a marker for lateral plate mesoderm⁵ and the earliest differentiation marker for endothelial cells and blood cells^{9,10}. We previously established an induction and purification system for Flk1⁺ cells from embryonic stem (ES) cells and showed that Flk1⁺ cells give rise to endothelial cells as well as blood cells *in vitro*^{11–13}. To elucidate the mechanisms underlying vascular development, we

attempted here to reconstitute the pathway required for blood vessel formation.

Undifferentiated ES cells (E-cadherin⁺) were cultured for four days on collagen IV coated dishes with 10% fetal calf serum (FCS) to induce Flk1⁺ cells^{11,12}. Flk1⁺/E-cadherin⁻ cells (at a purity of greater than 95%) obtained by flowcytometry sorting (Fig. 1a) did not express other endothelial cell (vascular endothelial cadherin (VEcad), platelet-endothelial cell adhesion molecule-1 (PECAM1),

CD34)¹¹ or mural cell (α -smooth muscle actin (SMA), PDGF- β receptor) markers (data not shown). After an additional four days of culturing of Flk1⁺ cells with 10% FCS, more than 95% of cells became positive for SMA (Fig. 1b). As unsorted cells generated heterogeneity in cell appearance with few SMA⁺ cells and PECAM1⁺ cells under the same conditions (Fig. 1c), this phenomenon appears specific to purified Flk1⁺ cells.

Addition of 50 ng ml⁻¹ VEGF (VEGF₁₆₅) to culture resulted in the appearance of PECAM1⁺ sheets of endothelial cells (Fig. 1d, n), which were also positive for VEcad, CD34, endoglin and acetyl-LDL incorporation (Fig. 1e–h). The remaining cells surrounding these sheets were SMA⁺ (Fig. 1d, n) and expressed other mural cell markers such as CGA7, a marker for differentiated vascular smooth muscle cells (VSMC)^{14,15} (Fig. 1i–l). Analysis by polymerase chain reaction with reverse transcription confirmed that other SMC

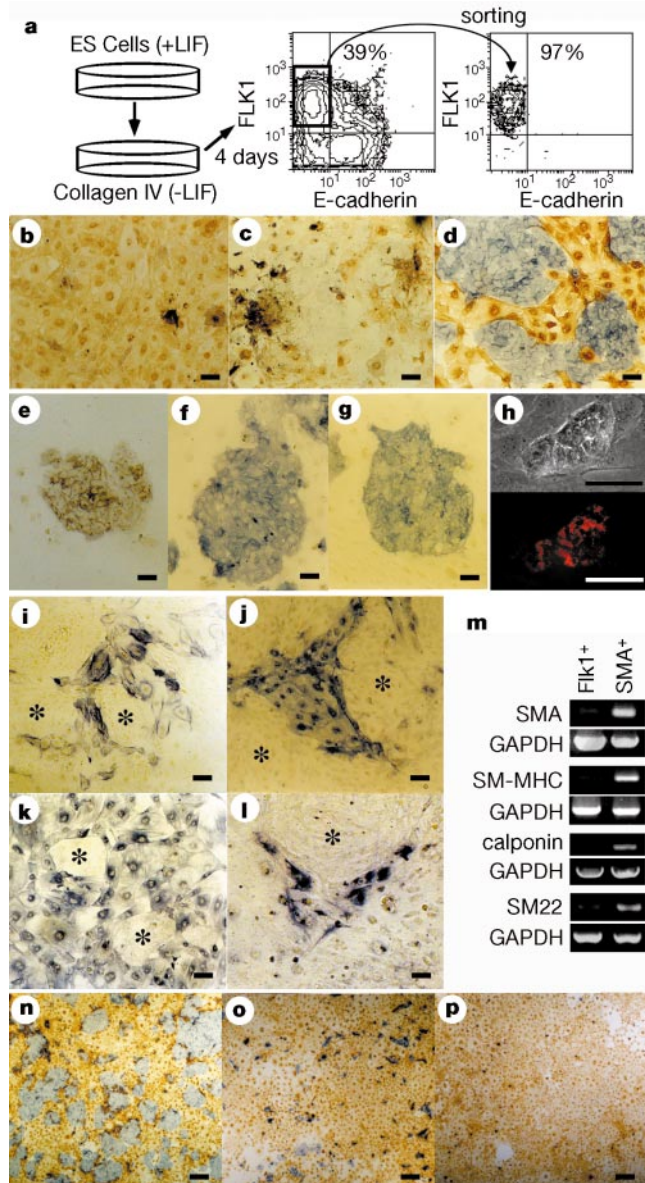


Figure 1 Differentiation of Flk1⁺ cells into endothelial and mural cells. **a**, Purification of Flk1⁺ cells from embryonic stem (ES) cells. Undifferentiated ES cells (E-cadherin⁺) cultured without LIF for four days induced Flk1 expression. Flk1⁺/E-cadherin⁻ population was purified by flowcytometry sorting. **b–d**, PECAM1 (purple) and SMA (1A4; brown) immunostaining. **b**, 10% FCS results in >95% SMA⁺ cells. **c**, Heterogenous cells are obtained in unsorted fractions. **d**, Flk1⁺ cells with 10% FCS and VEGF produce PECAM1⁺ sheets; expression of PECAM1 and SMA are mutually exclusive. PECAM1⁺ sheets were examined for VEcad (**e**), CD34 (**f**), endoglin (**g**) and acetyl-LDL incorporation (**h**). Immunostaining for mural cell markers: **i**, CGA7 (SMA in differentiated vascular SMC); **j**, PDGF- β receptor; **k**, smooth muscle tropomyosin; **l**, desmin. Asterisks, endothelial cell sheets negative for mural cell markers. **m**, RT-PCR analysis of mural cell markers in Flk1⁺ and SMA⁺ cells. **n–p**, PECAM1/SMA double immunostaining of Flk1⁺ cells with 50 ng ml⁻¹ VEGF₁₆₅ (**n**), VEGF₁₂ (**o**), PIGF (**p**). Scale bars: **b–h**, 100 μ m; **n–p**, 400 μ m.

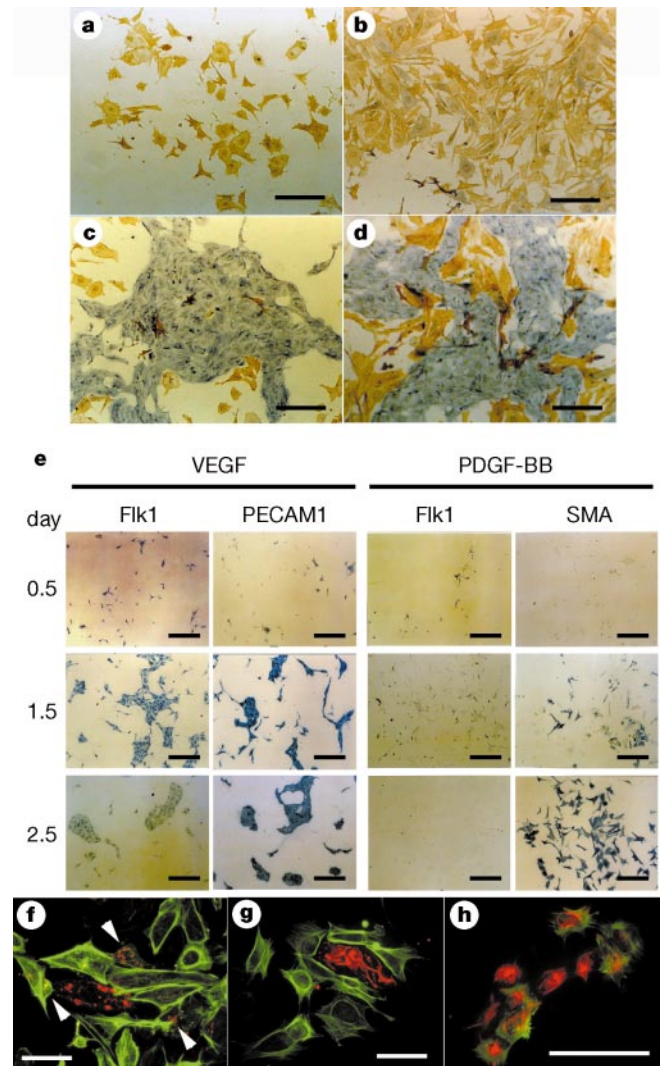


Figure 2 Flk1⁺ cell differentiation in serum-free culture. **a–d**, PECAM1 (purple)/SMA (1A4; brown) immunostaining (day 3). **a**, Vehicle-treated cells. **b**, PDGF-BB. **c**, VEGF. **d**, Simultaneous administration of VEGF and PDGF-BB. **e**, Time course of Flk1⁺ cell differentiation. VEGF maintains expression of Flk1 and PECAM1⁺ sheets from day 1.5. PDGF-BB-treated cells gradually lose Flk1 expression, SMA⁺ mural cells appear and dominate from day 1.5 to 2.5. **f**, Fluorescent immunostaining for Flk1 (red) and SMA (green) in 1.5-day PDGF-BB treatment. Arrowheads, Flk1⁺/SMA⁺ cells. **g**, VEcad (red)/SMA (green) double staining in 1.5-day PDGF-BB treatment. **h**, Flk1 (red)/SMA (green) double staining in 2.5-day VEGF and 10% FCS treatment. Flk1⁺/SMA⁺ cells are present. VEGF (50 ng ml⁻¹), PDGF-BB (10 ng ml⁻¹). Scale bars: **a–e**, 200 μ m; **f–h**, 100 μ m.

markers such as smooth muscle myosin heavy chain, calponin and SM22 α were upregulated in SMA⁺ cells (Fig. 1m), indicating that Flk1⁺ cells predominantly differentiate into either endothelial cells or mural cells under these culture conditions. As VEGF₁₆₅ was more efficient than VEGF₁₂₁ in endothelial cell induction (Fig. 1n, o) and placental growth factor (PIGF) had no activity (Fig. 1p), Flk1 and neuropilin, but not Flt1, are involved in this process¹⁶.

To define precisely the factors required for this differentiation, we tested the effects of various growth factors on Flk1⁺ cells cultured in

serum-free conditions. Some SMA⁺ cells were observed without addition of factors (Fig. 2a), indicating that de novo differentiation of mural cells could occur independently of exogenous growth factors. Administration of PDGF-BB but not PDGF-AA resulted in selective induction of SMA⁺ cells with spindle-like shapes resembling VSMC (Fig. 2b). This is consistent with previous reports indicating that PDGF-B/PDGF- β R signalling is required for vascular integrity through involvement in pericyte proliferation and recruitment to the vascular wall^{15,17–19}. In contrast, VEGF (50 ng ml⁻¹) administration induced predominantly PECAM1⁺ sheets (Fig. 2c). Both types of cell were increased by the simultaneous administration of VEGF and PDGF-BB (Fig. 2d), indicating that the induction of vascular cells from Flk1⁺ cells is probably regulated by a balanced combination of growth factors.

With VEGF treatment, Flk1 expression was maintained and Flk1⁺ cells formed PECAM1⁺ endothelial cell sheets after 1.5 days of incubation (Fig. 2e). In contrast, PDGF-BB-treated cells gradually lost Flk1 expression from 1.5 days and SMA⁺ mural cells appeared and became dominant from day 1.5 to 2.5 (Fig. 2e). Double fluorescent immunostaining of Flk1 and SMA in 1.5-day cultures of PDGF-BB-treated cells showed Flk1⁺/SMA⁺ cells as well as Flk1 and SMA single-positive cells (Fig. 2f). VEGF⁺ endothelial cells were distinct from SMA⁺ cells (Fig. 2g). Flk1⁺/SMA⁺ cells were also detected in the cultures with serum (Fig. 2h), unlike VEGF⁺/SMA⁺ cells. Although transdifferentiation of endothelial cells to SMA⁺ cells in chick embryo has been reported²⁰, VEGF⁺/SMA⁺ cells rarely appeared from Flk1⁺ cells under our culture conditions with or without serum. Thus, transdifferentiation of VEGF⁺ cells to SMA⁺ cells may rarely occur in our experimental system. These results indicate that VEGF is required for maintenance of Flk1 expression and differentiation to endothelial cells. In the absence of VEGF, Flk1 expression is lost and cells proliferate and differentiate into mural

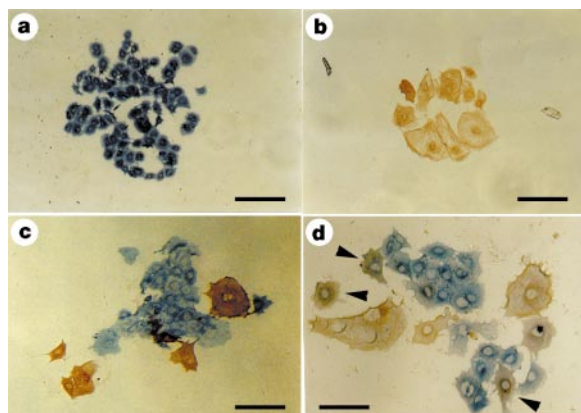


Figure 3 Single-cell deposition of Flk1⁺ cells cultivated with 10% FCS and 50 ng ml⁻¹ VEGF for four days and stained for PECAM1 (purple) and SMA (1A4; brown). Three colony types were observed. **a**, Pure endothelial cells (PECAM1⁺). **b**, Pure mural cells (SMA⁺). **c**, **d**, Mixed colonies consisting of PECAM1⁺ and SMA⁺ cells with some cells expressing both PECAM1 and SMA (arrowheads). Scale bars: 50 μ m.

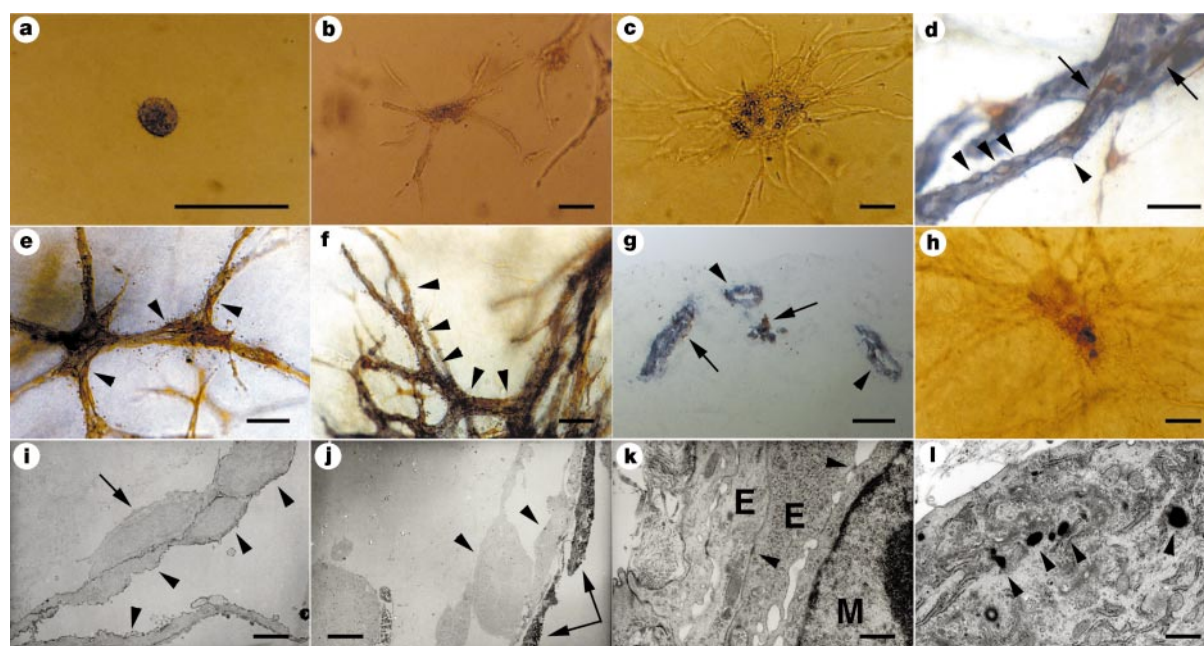


Figure 4 Vascular formation of Flk1⁺ cell aggregates in three-dimensional culture with 10% serum and 50 ng ml⁻¹ VEGF. **a–c**, Time course of tube formation. **a**, Flk1⁺ cell aggregate, day 0; **b**, day 3; **c**, day 5. **d–f**, In-gel double staining of PECAM1 (purple) and SMA (brown). **d**, PECAM1⁺ tube structure; endothelial cell nuclei (arrowheads) and attached SMA⁺ cells (arrows). **e**, Elongated SMA⁺ cells attached to tube with stellate processes (arrowheads). **f**, Tree-like structure with SMA⁺ cells (arrowheads). **g**, Cross-section reveals lumen with PECAM1⁺ endothelial cells (arrowheads) and attached SMA⁺ cells (arrows). **h**, Round cells in centre of aggregates are positive for blood cell markers;

antibody mixture for CD45 and Ter119 (brown). Immunoelectron microscopy for PECAM1 (**i**) and SMA (**j**). PECAM1⁺/SMA⁺ endothelial tube structure (arrowheads) with adhering PECAM1⁺/SMA⁺ mural cells (arrows). **k**, Junction formation in Flk1⁺ cell-derived structure. Desmosome-like structures (arrowheads) were observed in endothelial–endothelial and endothelial–mural cell junctions. E, endothelial cell; M, mural cell. **l**, Ultrastructure of endothelial cell revealing Weibel–Palade bodies (arrowheads). Scale bars: **a–c**, 100 μ m; **d–h**, 25 μ m; **i**, **j**, 5 μ m; **k**, **l**, 1 μ m.

cells through mainly PDGF-BB signalling. Although transforming growth factor- β (TGF- β) was reported to promote differentiation of mesenchymal cells to pericytes²¹, TGF- β inhibited growth of Flk1⁺ cells in our system (data not shown).

To determine whether Flk1⁺ cells contain common progenitors for endothelial and mural cells, we performed single cell deposition. Three types of colony emerged from single Flk1⁺ cells: pure endothelial cells (PECAM1⁺), pure mural cells (SMA⁺) and mixtures of both (Fig. 3a–d). Mixed colonies also contained PECAM1⁺/SMA⁺ cells (Fig. 3d, arrowheads). The frequencies of endothelial cell, mural cell and mixed colonies in the limiting dilution assay were around 27%, 43% and 30%, respectively. Plating efficiency was 14%. Similarly, Flk1⁺ cells (Flk1⁺/VEcad⁺) obtained from embryonic day (E) 8.5 embryos (85–90% purity) gave rise to all three colony types: endothelial cells (11%), mural cells (80%) and mixed (9%) (data not shown), indicating that Flk1⁺ cells from ES cells as well as embryos possess the potential to serve as common 'vascular progenitor cells' (VPC).

We next tested whether Flk1⁺ cells can form a vascular structure *in vitro*. Aggregates of around several hundred Flk1⁺ cells from ES cells were cultivated in collagen I gels with 10% FCS and 50 ng ml⁻¹ VEGF. Cells migrated out from the aggregates and formed tube structures within 3–5 days (Fig. 4a–c; see Supplementary Information). Double immunostaining revealed that these structures consisted of PECAM1⁺ endothelial tubes associated with SMA⁺ cells, which attached to endothelial tubes with stellate processes (Fig. 4d, e), a feature compatible with that of pericytes in capillary vessels. Occasional formation of vascular tree-like structures with massive investment of SMA⁺ cells was observed (Fig. 4f). Cross-sections showed a true lumen with attached SMA⁺ cells (Fig. 4g). In addition, spherical cells within the lumen were positive for the blood cell markers CD45 and Ter119, indicating generation of blood cells (Fig. 4h). This *in vitro* culture system of Flk1⁺ cell aggregates appears to mimic vascular development from yolk sac blood islands.

Immunoelectron microscopy suggested that the tube formation occurred by direct contact of PECAM1⁺ and SMA⁺ cells (Fig. 4i, j). Desmosome-like structures were formed in both endothelial–endothelial and endothelial–mural cell junctions (Fig. 4k), indicating direct interaction of these vascular cells. Relatively long adhesive structures were formed between endothelial cells, whereas endothelial and mural cells exhibited patchy formation of desmosome-like junctions²². Endothelial cells contained rough endoplasmic reticulum (RER), Golgi apparatus, microfilaments and sometimes round or rod-shaped electron-dense Weibel–Palade bodies (Fig. 4l)²³. Pericytes possessed thin stellate processes, abundant microfilaments, microtubules, RER and Golgi apparatus, but not Weibel–Palade bodies. There were a few collagen fibres adjacent to the pericytes, but we could not confirm clear basement membranes or external laminae. These results indicate that endothelial and mural cells that differentiate from Flk1⁺ cells interact with each other to form mature vessel-like structures *in vitro*.

To analyse the *in vivo* differentiation potential of Flk1⁺ cells, we induced LacZ-expressing Flk1⁺ cells from CCE/nLacZ (see Methods), injected them intracardially into stage 16–17 chick embryos and traced them by whole-mount staining. Most LacZ signals appeared along vessels in the head, yolk sac, heart and intersomitic region, in some cases forming a network structure 2–3 days after injection (Fig. 5a, b). Immunohistochemical staining of sections demonstrated incorporation of donor cells as both endothelial cells (PECAM1⁺/nLacZ⁺) (Fig. 5c) and mural cells (SMA⁺/nLacZ⁺) (Fig. 5d). This indicates that Flk1⁺ cells differentiate to both types of cell *in vivo* and contribute to the vascular component.

We thus showed that ES-derived Flk1⁺ cells give rise to two major vascular cell types (endothelial cells and mural cells) *in vitro* and *in vivo*. We started from a purified population of Flk1⁺ cells that allowed the specific differentiation of vascular cells and facilitated

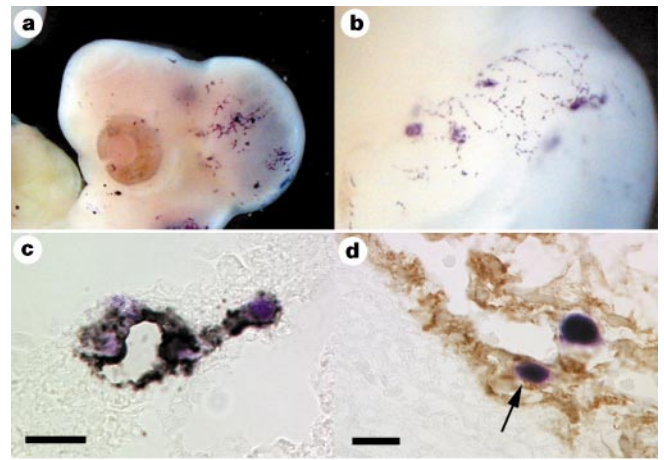


Figure 5 Contribution of Flk1⁺ cells to vascular formation *in vivo*. **a, b**, Whole-mount β-gal staining of chick embryos. **a**, 2 days; **b**, 3 days after injection of Flk1⁺ cells (CCE/nLacZ). LacZ⁺ nuclei formed network structure along vasculature. **c, d**, Immunostaining of sectioned LacZ-stained embryos for PECAM1 (grey) and SMA (1A4; brown). Anti-PECAM1 antibody is specific to murine endothelial cells. 1A4 reacts with both murine and chick SMA. **c**, Incorporation of PECAM1⁺/nLacZ⁺ cells into vasculature in yolk sac. Donor cells formed a true lumen. **d**, Incorporation of SMA⁺/nLacZ⁺ cells in head vascular wall (arrow). Blood cells are observed in the lumen. Scale bars: 10 μm.

analyses of vascular development. Our study gives new insight into the vascular developmental pathway that originates in Flk1⁺ cells (see Supplementary Information). Observations that other growth factors such as basic fibroblast growth factor induced cells which were negative for both PECAM1 and SMA (data not shown) indicate that Flk1⁺ cells can also differentiate into lineages other than vascular. Further roles of Flk1⁺ cells in embryogenesis should be explored and might provide additional potential of Flk1⁺ cells for tissue engineering beyond the vascular system. □

Methods

Cell culture and sorting

Maintenance, differentiation, culture and cell sorting of CCE ES cells (gift from M. J. Evans) were as described¹². We plated 2×10^4 Flk1⁺ cells per well on collagen IV (Col. IV)-coated 24-well dishes (Becton Dickinson) in differentiation medium¹². For serum-free culture, cells were incubated in SFO3 (Sanko Junyaku) as described²⁴. Single sorted cells were plated into individual wells of Col IV-coated 96-well dishes¹². In limiting dilution assay, 200–500 Flk1⁺ cells were plated per plate. Before three-dimensional culture, cells (4×10^5 cells per ml) were incubated in differentiation medium¹² containing 50 ng ml⁻¹ VEGF on uncoated petri dishes for 12 h to induce aggregation. Aggregates were resuspended in 2× differentiation medium and mixed with an isovolume of collagen I-A gel (3 mg ml⁻¹) (Nitta Gelatin). We plated 250 μl of this mixture onto a lucent insert disc, Cell Disk (Sumitomo Bakelite), in 24-well dishes. After 15 min at 37 °C to allow polymerization, we added 500 μl differentiation medium with VEGF (final 50 ng ml⁻¹). Human VEGF₁₆₅, VEGF₁₂₁ and PlGF were purchased from R&D Systems Inc., and human PDGF-AA and PDGF-BB from GIBCO BRL.

Monoclonal antibodies

Monoclonal antibodies for murine E-cadherin (ECCD2), Flk1 (AVAS12) and VEcad (VECD1) as described^{25–27} and murine PDGF-βR (APB5) were prepared and labelled in our laboratory. Monoclonal antibodies for murine PECAM1 (Mec13.3), VEcad (11D4.1) for immunohistochemistry, Ter119 and CD45 were purchased from Pharmingen; those for SMA (1A4, CGA7) from NeoMarkers and ENZO Diagnostics, respectively; and those for desmin, tropomyosin and endoglin from DAKO, Sigma and Southern Biotechnology Associates, respectively.

Immunohistochemistry

Single staining of culture cells on dishes was as described¹¹. For double staining, cells were incubated with a mixture of Mec13.3 and 1A4 followed by ALP-conjugated anti-rat IgG (H+L) (Jackson ImmunoResearch Lab. Inc.) and HRP-conjugated anti-mouse IgG (BIO SOURCE International). For three-dimensional cultures, gel blocks on discs were fixed in 4% paraformaldehyde for 5 min at room temperature, then processed for whole-mount immunohistochemistry²⁸. Stained blocks were frozen in OCT compound (Miles Inc.) and sectioned by Cryostat (MICROM) at 5–7 μm thickness.

Electron microscopy

Electron microscopy and immunoelectron microscopy were performed as described^{29,30}. For electron microscopy, cell samples were fixed in 1.5% glutaraldehyde for 2 h. For immunoelectron microscopy, cells were fixed with 0.1% glutaraldehyde for 10 min.

Reverse transcription-polymerase chain reaction (RT-PCR)

After four days culture of Flk1⁺ cells with 10% FCS, cells (Fig. 1b) were used as a source of SMA⁺ RNA. Total RNA was prepared with TRIzol reagent (Life Technologies, Inc.), reverse transcribed by oligo (dT) priming and PowerScript-Reverse Transcriptase (CLONTECH Laboratories), and PCR was performed with the following primers: SMA forward: 5'-ACGGCCGCTCTCTTCTCTC-3', reverse 5'-GCCAGCTTCGTCGATTCC-3', smooth muscle myosin heavy chain forward: 5'-GACAACTCTCTCGCTTTGG-3', reverse: 5'-GCTCTCCAAAAGCAGGTAC-3', h1-calponin forward: 5'-GATACGAATTCAGAGGGTGCAGACGGAGGCTC-3', reverse: 5'-GATACAAGCTTTCAATCCACTCTCTCAGCTCC-3', SM22 α forward: 5'-CGACTCCAAAATTGAGAAGA-3', reverse: 5'-CTGTGTCTGCCCATTTGAAG-3'.

Chick/mouse chimaeric assay

CCE/nLacZ cells were generated by cotransfection of elongation factor 1 promoter-driven LacZ gene with nuclear localization signal (T. Kunisada) and murine phosphoglycerate kinase 1 promoter-driven puromycin resistant gene constructs and selection by 2 μ g ml⁻¹ puromycin. We injected 1–2 $\times$ 10⁵ Flk1⁺ cells in 2–4 μ l of phosphate buffered saline (PBS) into hearts of stage 16–17 chick embryos with glass needles. Embryos were killed 2–3 days after injection and fixed with 2% paraformaldehyde at 4 °C for 10 min. A X-gal analogue, magenta gal (Nakarai), was used as substrate for whole-mount staining; stained embryos were frozen in OCT, cryosectioned at 6–7 μ m and stained for PECAM1 or SMA. PECAM1 staining was performed using TSA Indirect, tyramide signal amplification reagent (NEN Life Science Products).

Received 26 June; accepted 23 August 2000.

- Suri, C. *et al.* Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).
- Folkman, J. & D'Amore, P. A. Blood vessel formation: What is its molecular basis? *Cell* **87**, 1153–1155 (1996).
- Darland, D. C. & D'Amore, P. A. Blood vessel maturation: vascular development comes of age. *J. Clin. Invest.* **103**, 157–158 (1999).
- Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
- Yamaguchi, T. P., Dumont, D. J., Conlon, R. A., Breitman, M. L. & Rossant, J. Flk1, a flt-related receptor tyrosine kinase is an early marker for endothelial precursors. *Development* **118**, 489–498 (1993).
- Topouzis, S. & Majesky, M. W. Smooth muscle lineage diversity in the chick embryo—Two types of aortic smooth muscle cell differ in growth and receptor-mediated transcriptional responses to transforming growth factor- β . *Dev. Biol.* **178**, 430–445 (1996).
- Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P. & Sucov, H. M. Fate of the mammalian cardiac neural crest. *Development* **127**, 1607–1616 (2000).
- Mikawa, T. & Gourdie, R. G. Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with in growth of the epicardial organ. *Dev. Biol.* **174**, 221–232 (1996).
- Shalaby, F., Rosant, J., Yamaguchi, T. P., Gertsenstein, M. & Wu, X. F. Failure of blood-island formation and vasculogenesis in Flk1-deficient mice. *Nature* **376**, 62–66 (1995).
- Eichmann, A. *et al.* Ligand-dependent of the endothelial and hematopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor 2. *Proc. Natl Acad. Sci. USA* **94**, 5141–5146 (1997).
- Hirashima, M., Kataoka, H., Nishikawa, S., Matsuyoshi, N. & Nishikawa, S. I. Maturation of embryonic stem cells into endothelial cells in an *in vitro* model of vasculogenesis. *Blood* **93**, 1253–1263 (1999).
- Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N. & Kodama, H. Progressive lineage analysis by cell sorting and culture identifies FLK1⁺VE-cadherin⁺ cells at a diverging point of endothelial cells and hemopoietic lineages. *Development* **125**, 1747–1757 (1998).
- Ogawa, M. *et al.* Expression of α 4-integrin defines the earliest precursor of hematopoietic cell lineage diverged from endothelial cells. *Blood* **93**, 1168–1177 (1999).
- Ueda, M., Becker, A. E., Naruko, T. & Kojima, A. Smooth muscle cell de-differentiation is a fundamental change preceding wound healing after percutaneous transluminal coronary angioplasty in humans. *Coron. Artery Dis.* **6**, 71–81 (1995).
- Hellström, M., Kalin, M., Lindahl, P., Abramson, A. & Betsholtz, C. Role of PDGF-B and PDGFR- β in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–3055 (1999).
- Soker, S., Takashima, S., Miao, H. Q., Neufeld, G. & Klagsbrun, M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* **92**, 735–745 (1998).
- Lindahl, P., Johansson, B., Leveén, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277**, 242–245 (1997).
- Hirschi, K. K., Rohovsky, S. A., Beck, L. H., Smith, S. R. & D'Amore, P. A. Endothelial cells modulate the proliferation of mural cells precursors via platelet-derived growth factor-BB and heterotypic cell contact. *Circ. Res.* **84**, 298–305 (1999).
- Benjamin, L. E., Hemo, I. & Keshet, E. A plasticity window for blood vessel remodelling is defined by pericyte coverage of the preformed endothelial network and is regulated by PDGF-B and VEGF. *Development* **125**, 1591–1598 (1998).
- DeRuiter, M. C. *et al.* Embryonic endothelial cells transdifferentiate into mesenchymal cells expressing smooth muscle actins in vivo and in vitro. *Circ. Res.* **80**, 444–451 (1997).
- Hirschi, K. K., Rohovsky, S. A. & D'Amore, P. A. PDGF, TGF- β , and heterotypic cell-cell interactions mediate endothelial cell-induced recruitment of 10T1/2 cells and their differentiation to a

smooth muscle fate. *J. Cell. Biol.* **141**, 805–814 (1998).

- Cuevas, P. *et al.* Pericyte endothelial gap junctions in human cerebral capillaries. *Anat. Embryol.* **170**, 155–159 (1984).
- Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
- Takakura, N., Kodama, H., Nishikawa, S. & Nishikawa, S. I. Preferential proliferation of murine colony-forming units in culture in a chemically defined condition with a macrophage colony-stimulating factor-negative stromal cell clone. *J. Exp. Med.* **184**, 2301–2309 (1996).
- Shirayoshi, Y., Nose, A., Iwasaki, K. & Takeichi, M. N-linked oligosaccharides are not involved in the function of a cell-cell binding glycoprotein E-cadherin. *Cell Struct. Funct.* **11**, 245–252 (1986).
- Kataoka, H. *et al.* Expressions of PDGF-receptor alpha, c-kit, and FLK1 genes clustering in mouse chromosome 5 define distinct subsets of nascent mesodermal cells. *Dev. Growth Differ.* **39**, 729–740 (1997).
- Matsuyoshi, N. *et al.* *In vivo* evidence of the critical role of cadherin-5 in murine vascular integrity. *Proc. Assoc. Am. Physicians* **109**, 362–371 (1997).
- Yoshida, H. *et al.* IL-7 receptor α CD3⁺ cells in the embryonic intestine induces the organizing center of Peyer's patches. *Int. Immunol.* **11**, 643–655 (1999).
- Isobe, S., Chen, S. T., Nakane, P. K. & Brown, W. R. Studies on translocation of immunoglobulins across intestinal epithelium. I. Improvements to study the peroxidase-labeled antibody method for application to study of human intestinal mucosa. *Acta Histochem. Cytochem.* **10**, 161–171 (1977).
- Yamamoto, T. *et al.* Repopulation of murine Kupffer cells after intravenous administration of liposome-encapsulated dichloromethylene diphosphonate. *Am. J. Pathol.* **149**, 1271–1286 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. J. Evans for CCE ES cells, N. Matsuyoshi for the hybridoma, A. Nagafuchi for antibodies, T. Kunisada for LacZ construct, R. Yu for critical reading of the manuscript, and many of our colleagues for suggestions and discussion. This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan, the Ministry of Health and Welfare of Japan (S.I.N.), Japanese Society for the Promotion of Science "Research for the Future" Program (H.I., M.O. and S.I.N.), Japan Tobacco Foundation, Japan Hearth Foundation & Pfizer Pharmaceuticals Grant for Research on Coronary Artery Disease and Tanabe Medical Frontier Conference. J.Y. is a recipient of the Research Fellowship grant of the Japan Society for the Promotion of Science for Young Scientists.

Correspondence and requests for materials should be addressed to J.Y. (e-mail: juny@kuhp.kyoto-u.ac.jp).

Bidirectional control of airway responsiveness by endogenous cannabinoids

A. Calignano[†], I. Kátóná[†], F. Désarnaud[‡], A. Giuffrida[‡], G. La Rana^{*}, K. Mackie[§], T. F. Freund[†] & D. Piomelli[†]

^{*} Department of Pharmacology, University of Naples, Naples 80131, Italy

[†] Institute of Experimental Medicine, Budapest 1450, Hungary

[‡] Department of Pharmacology, University of California, Irvine, California 92697-4625, USA

[§] Department of Anesthesiology, University of Washington, Seattle, Washington 98195-6540, USA

Smoking marijuana or administration of its main active constituent, Δ^9 -tetrahydrocannabinol (Δ^9 -THC), may exert potent dilating effects on human airways^{1–4}. But the physiological significance of this observation and its potential therapeutic value are obscured by the fact that some asthmatic patients respond to these compounds with a paradoxical bronchospasm^{3,5}. The mechanisms underlying these contrasting responses remain unresolved. Here we show that the endogenous cannabinoid anandamide exerts dual effects on bronchial responsiveness in rodents: it strongly inhibits bronchospasm and cough evoked by the chemical irritant, capsaicin, but causes bronchospasm when the constricting tone exerted by the vagus nerve is removed. Both effects are mediated through peripheral CB1 cannabinoid receptors found on axon terminals of airway nerves. Biochemical

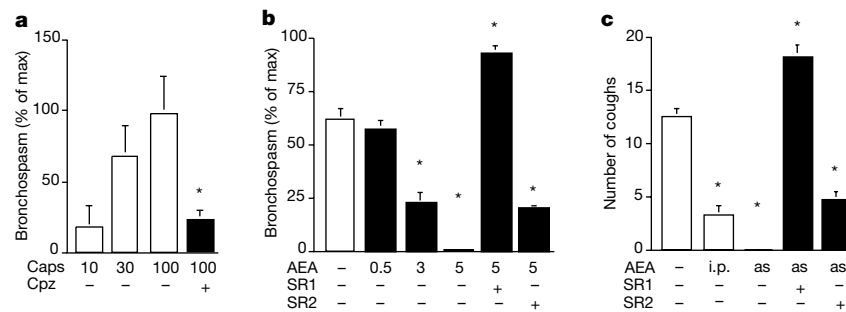


Figure 1 Anandamide inhibits bronchospasm and coughing in guinea-pigs by activating peripheral CB1 receptors. **a**, Constricting effect of capsaicin (Caps, μg per kg i.v.) on guinea pig bronchial smooth muscle and its antagonism by capsazepine (Cpz, 0.2 mg per kg i.v.). **b**, Inhibitory effects of anandamide (AEA, mg per kg i.v.) on capsaicin-evoked (30 μg per kg) bronchospasm in the absence or presence of the CB1 antagonist

SR141716A (SR1, 0.5 mg per kg i.v.) or the CB2 antagonist SR144528 (SR2, 0.3 mg per kg i.v.). **c**, Inhibition of capsaicin-evoked cough by anandamide in the absence or presence of SR141716A (0.5 mg per kg i.v.) or SR144528 (0.3 mg per kg i.v.). Asterisk, $P < 0.01$ ($n = 3$ for each condition).

analyses indicate that anandamide is synthesized in lung tissue on calcium-ion stimulation, suggesting that locally generated anandamide participates in the intrinsic control of airway responsiveness. In support of this conclusion, the CB1 antagonist SR141716A enhances capsaicin-evoked bronchospasm and cough. Our results may account for the contrasting bronchial actions of cannabis-like drugs in humans, and provide a framework for the development of more selective cannabinoid-based agents for the treatment of respiratory pathologies.

To explore the functional roles of the cannabinoid system in airway physiology, we investigated the effects of the endogenous cannabinoid, anandamide^{6,7}, on the responsiveness of bronchial smooth muscle. Administration of capsaicin, the pungent component of chilli pepper⁸, produces a potent bronchoconstriction in anaesthetized guinea pigs (Fig. 1a) or rats (in μg per animal, intratracheal: 10, $41 \pm 6\%$ of maximal bronchospasm; 30, $55 \pm 12\%$; 100, $81 \pm 19\%$; means $\pm$ s.e.m., $n = 3$). This response, which has been extensively studied both in animals and humans, is thought to result from the activation of capsaicin (or 'vanilloid') receptors on sensory C-fibres⁹. Accordingly, the constricting effects of capsaicin were blocked by the vanilloid antagonist capsazepine (Fig. 1a). When administered systemically before capsaicin, anandamide produced a dose-dependent attenuation of bronchospasm in guinea-pigs (Fig. 1b) and rats (capsaicin, 10 μg per animal, intratracheal: $37.2 \pm 4.2\%$ of maximal bronchospasm; capsaicin after anandamide, 1 mg per kg intravenous (i.v.), $14 \pm 7\%$, $n = 3$). This effect was completely reversed by the selective CB1 cannabinoid antagonist SR141716A (ref. 10) but only slightly reduced with a maximal dose of the CB2 antagonist SR144528 (ref. 11) (Fig. 1b and data not shown). Palmitylethanolamide, a structural analogue of anandamide that inhibits nociception in mice, but does not interact with CB1 receptors¹², was ineffective at alleviating capsaicin-evoked bronchospasm (data not shown). When given as an aerosol to conscious guinea-pigs, capsaicin stimulates C-fibre activity in the upper respiratory tract and triggers coughing^{13,14}. Systemic or aerosolized anandamide reduced capsaicin-evoked coughing, an effect that was cancelled by CB1 receptor blockade (Fig. 1c).

Anandamide had no direct bronchomotor action, except at the highest dose tested (5 mg per kg), at which the compound elicited a small bronchoconstriction ($11.8 \pm 5.9\%$ of maximal; mean $\pm$ s.e.m., $n = 5$). To investigate this response further, we examined the effects of anandamide in anesthetized rodents that were deprived of the bronchoconstricting tone conferred by the vagus nerve⁹. After vagotomy and atropine administration, which eliminate vagal influences, systemic application of anandamide produced a dose-dependent bronchoconstriction in guinea pigs (Fig. 2a) and rats (in mg per kg i.v.: 1, $0 \pm 0\%$ of maximal bronch-

ospasm; 3, $12 \pm 1.7\%$, 5, $18.3 \pm 1.2\%$; $n = 3$). Similar effects were observed when anandamide was injected into the guinea-pig bronchi through a tracheal catheter (Fig. 2b), or applied to isolated strips of guinea-pig lung parenchyma (Fig. 2c–e). The slow onset of the anandamide response in strips of guinea-pig lung is consistent with results obtained in other isolated tissues⁶; the low potency of anandamide in this preparation may be accounted for by limited tissue penetration and/or rapid inactivation. In agreement with this possibility, the anandamide transport inhibitor AM404 enhanced anandamide-evoked contractions in isolated strips of guinea-pig lung (anandamide, 50 μM , 0.336 ± 0.07 dyne mg^{-1} of tissue; anandamide plus AM404, 28 μM , 0.638 ± 0.06 dyne mg^{-1} of tissue; $P < 0.05$, $n = 6$). The CB1 antagonist SR141716A blocked

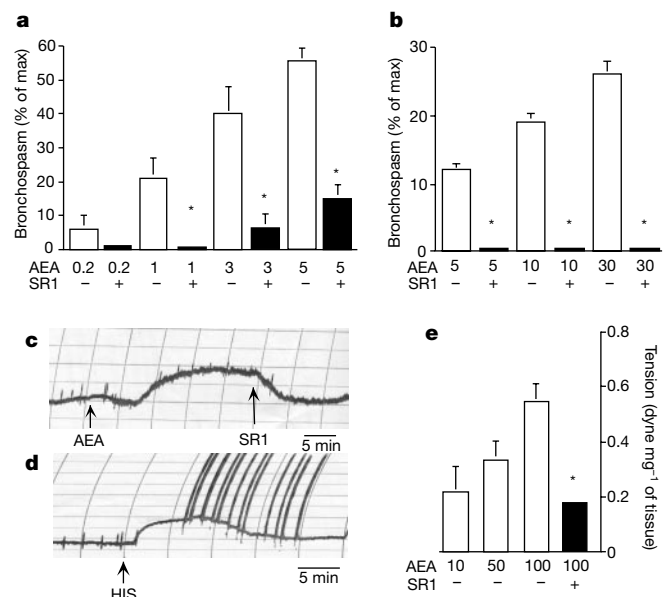


Figure 2 Anandamide causes bronchoconstriction in vagotomized, atropine-treated guinea pigs by activating peripheral CB1 receptors. **a**, Effects of anandamide (AEA, mg per kg i.v.) in the presence or absence of the CB1 antagonist SR141716A (SR1, 0.2 mg per kg i.v.) ($n = 6$ for each condition). **b**, Effects of anandamide (5–30 μg per animal, intratracheal) in the absence or presence of SR141716A (SR1, 0.3 mg per kg i.v.) ($n = 6$). **c**, Representative tracing illustrating the effect of anandamide (100 μM) on muscle tension in guinea pig parenchyma strips and its reversal by the CB1 antagonist SR141716A (1 μM). **d**, Response of the same lung strip to histamine (His, 10 μM). **e**, Contractions of lung parenchyma by anandamide (μM) and antagonism by SR141716A (1 μM) ($n = 6$). Asterisk, $P < 0.01$.

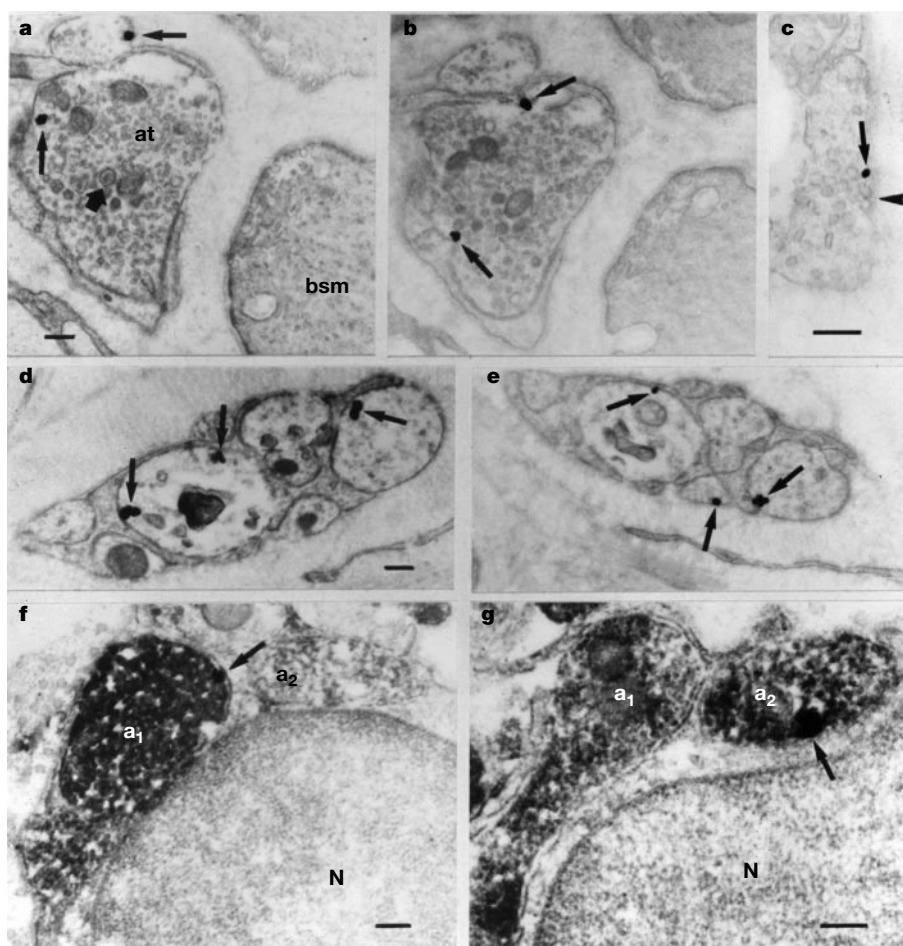


Figure 3 Localization of CB1 receptors on axon terminals and preterminal segments in rat lungs. **a, b**, CB1 receptor labelling is indicated by thin arrows on sections of bronchiole axon terminals. Axon terminals (at) containing small electron-translucent and large dense-core vesicles (thick arrow) are embedded in the collagen matrix and are surrounded by bronchial smooth muscle cells (bsm). **c**, CB1 receptor labelling close to

putative neurotransmitter release sites (arrowhead). **d, e**, Several axon terminals packed together into glial capsules in the adventitia. **f, g**, Co-localization of CB1 receptor and NPY. CB1 receptor labelling appears on axon terminal membranes (a_1 and a_2) displaying electron-dense NPY immunoreactivity. N, nucleus of a putative Schwann cell. Scale bars, 0.2 μm (scales in **b** and **e** are the same as in **a** and **d**, respectively).

anandamide bronchoconstriction *in vivo* and *in vitro* (Fig. 2a–e), whereas the CB2 antagonist SR144528 had no effect (data not shown). The cannabinoid agonist HU210 was also potent at eliciting guinea pig bronchial muscle constriction after tracheal administration (in μg per animal: $10.0 \pm 0.6\%$ of maximal bronchoconstriction; 1, $30 \pm 1.2\%$; 10, $60 \pm 2.2\%$; 30, 100%; $n = 6$). Anandamide has been claimed to activate vanilloid receptors¹⁵. However, the vanilloid antagonist capsazepine had no effect on anandamide-evoked bronchoconstriction at a dose that completely prevented the capsaicin response (0.2 mg per kg *i.v.*) (data not shown). These results indicate that removing the vagal excitatory tone unmasked a bronchoconstricting activity of anandamide mediated through CB1 receptors.

The ability of anandamide to influence bronchial muscle contractility after local administration suggests that this compound may exert its effects by activating CB1 receptors located within the airways. To test this possibility, we examined the ultrastructural localization of CB1 receptors in rat lungs by electron microscopy, using an antibody directed against the intracellular carboxy terminus of the CB1 receptor protein. Immunogold staining revealed that CB1 receptors are present on nerve fibres distributed among bronchial and bronchiolar smooth muscle cells (Fig. 3a–c), or between the longitudinal and circular smooth muscle layers, where several axons are packed together into glial capsules (Fig. 3d, e). All bundles contained at least one CB1-receptor-

positive axon. Detailed evaluation (20 bundles consisting of 91 axons followed through at least 25 consecutive sections) revealed that 36% of the axons were labelled with the CB1 receptor antibody. The gold particles that labelled CB1 receptors were attached to the inner surface of the axon plasma membrane, either at the release site or in the preterminal segments. This is consistent with the fact that the antibody we used recognizes the intracellular C terminus of the CB1 receptor protein. Axon terminals bearing CB1 receptor immunoreactivity were in close proximity to smooth muscle cells (0.2–0.5 μm), and contained a large number of small agranular vesicles, along with few dense-core vesicles (Fig. 3a, b). In some cases, CB1 receptor immunoreactivity was adjacent to clusters of vesicles accumulated at the plasma membrane, which most probably represent neurotransmitter release sites (Fig. 3c). Next, to determine whether CB1 receptors are localized on noradrenaline-containing and/or non-noradrenaline-containing fibres, we used a combination of immunogold staining for CB1 receptors, and immunoperoxidase staining for neuropeptide Y (NPY), a co-transmitter in sympathetic neurons⁹. We found that 63% of NPY-bearing axons were also CB1-receptor-positive (Fig. 3f, g). Notably, however, extensive labelling was observed on many NPY-negative axons (data not shown), suggesting that both noradrenaline-containing and/or non-noradrenaline-containing nerves may express CB1 receptors.

The finding that CB1 receptors are found predominantly, if not

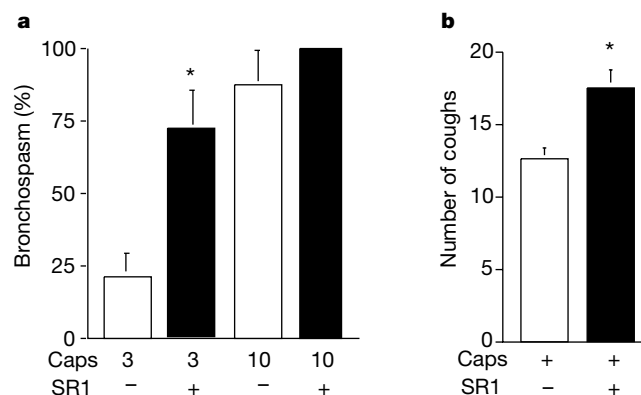


Figure 4 Intrinsic effects of the CB1 antagonist SR141716A on capsaicin-evoked bronchospasm and cough. **a**, Bronchoconstricting effects of capsaicin (μg per animal, intratracheal) in the absence or presence of the CB1 antagonist SR141716A (0.2 mg per kg i.v.). **b**, Tussigenic effects of capsaicin (0.3 mM, 4-min aerosol) in the absence or presence of SR141716A (0.2 mg per kg i.v.) ($n = 6$ for each condition). Asterisk, $P < 0.05$.

exclusively, on axon terminals of airway nerves indicates that anandamide may regulate bronchial smooth muscle tone through a prejunctional mechanism. Inhibition of excitatory neurotransmission in the airways may provide a parsimonious explanation for the ability of anandamide to oppose capsaicin-evoked bronchospasm and cough. This interpretation is further supported by the ability of anandamide and other cannabinoid agonists to inhibit neurotransmitter release in both peripheral tissues^{16–18} and the central nervous system¹⁹. The mechanism underlying the constricting actions of anandamide in animals lacking acetylcholine-mediated control is currently unknown. One possibility, which is consistent with the co-localization of CB1 receptors with NPY, is that anandamide inhibits the release of a bronchodilating mediator. Alternatively, anandamide may interact with CB1 receptors on smooth muscle. Our failure to detect CB1 receptor immunoreactivity in lung smooth muscle may have been caused by insufficient sensitivity of our technique, or by the presence in smooth muscle of a receptor variant that is not recognized by our antibody. Interestingly, northern blot analyses suggest that alveolar type II cells in the lung may express two different CB1 receptor messenger RNA species²⁰.

To test the possibility that endogenous cannabinoids regulate airway responsiveness, we determined the intrinsic effects of CB1 and CB2 antagonists on bronchospasm and cough in guinea-pigs. CB1 receptor blockade with SR141716A had no bronchomotor consequences *per se*, but significantly enhanced the bronchoconstriction and coughing evoked by capsaicin administered through a tracheal catheter (Fig. 4a, b), even in the presence of anandamide (Fig. 1b, c). This response did not depend on the capsaicin administration route, as it was also seen after i.v. injection of the drug (30 mg per kg capsaicin alone, $55.3 \pm 8.2\%$ of maximal bronchospasm; capsaicin after SR141716 (0.5 mg per kg i.v.), $92.3 \pm 3.4\%$; $P < 0.05$, $n = 3$). The CB2 antagonist SR144528 had no effect on capsaicin-induced bronchoconstriction and cough (data not shown).

Although the bronchomotor actions of the CB1 antagonist may be accounted for by its inverse agonist properties²¹, evidence suggests that this drug acted by opposing an ongoing cannabinoid modulation. First, the lack of effect seen with the CB1 antagonist in the absence of capsaicin is incompatible with an inverse agonist behaviour. Second, analyses by high-performance liquid chromatography (HPLC) coupled to positive-ionization electrospray mass spectrometry (MS) revealed that anandamide is synthesized in rat lung tissue through a Ca^{2+} -ion-activated mechanism (Fig. 5). Rat

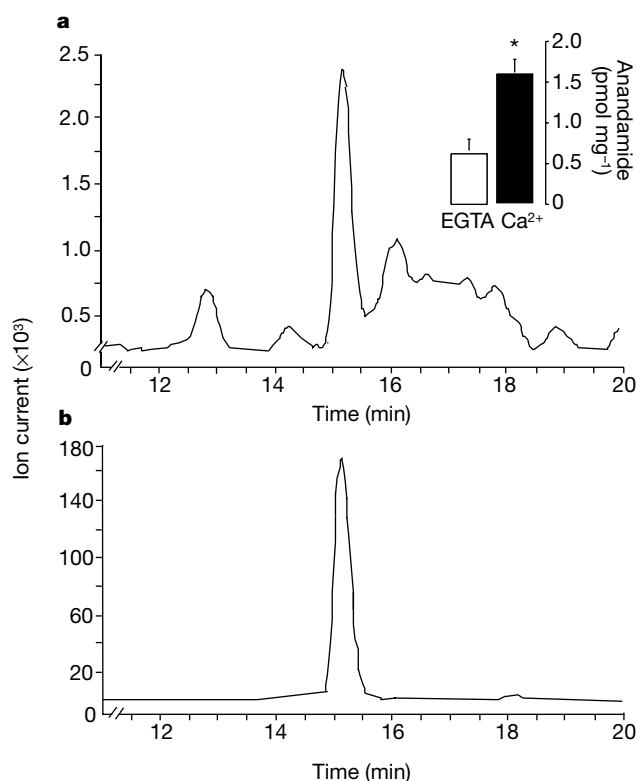


Figure 5 Ca^{2+} -dependent biosynthesis of anandamide in rat lung tissue. Representative HPLC/MS tracings for selected ions characteristic of endogenous anandamide (**a**, mass-to-charge ratio $m/z = 370$, an adduct with Na^+ , $[\text{M} + \text{Na}]^+$) and synthetic [$^2\text{H}_4$]anandamide (**b**, $m/z = 374$, $[\text{M} + \text{Na}]^+$), which was added to the samples as an internal standard. The inset in **a** illustrates the effects of EGTA (1 mM) or Ca^{2+} (3 mM) on anandamide biosynthesis in rat lung membranes. Ca^{2+} significantly stimulated anandamide formation (mean $\pm$ s.e.m.; asterisk, $P < 0.05$, $n = 4$).

lung membranes produced on average 0.6 ± 0.2 pmol of anandamide per mg of protein in the presence of the Ca^{2+} chelator EGTA (1 mM); and 1.6 ± 0.2 pmol of anandamide per mg of protein in the presence of Ca^{2+} (3 mM) (mean $\pm$ s.e.m., $n = 4$; $P < 0.05$ between EGTA and Ca^{2+} ; Student's *t*-test) (Fig. 5, inset). Guinea-pig lung membranes produced 0.9 ± 0.3 pmol of anandamide per mg of protein in the presence of EGTA; and 8.8 ± 1.2 pmol of anandamide per mg of protein in the presence of Ca^{2+} ($n = 4$; $P < 0.0001$; Student's *t*-test).

Anandamide is thought to originate from the enzymatic cleavage of *N*-arachidonyl phosphatidylethanolamine (NAPE), the biosynthesis of which is catalysed by a Ca^{2+} -dependent *N*-acyltransferase activity^{7,22,23}. Using negative ionization electrospray HPLC/MS, we identified two molecular species of NAPE in lipid extracts of rat lung membranes: alk-1-palmitoonyl-2-arachidonyl-*sn*-glycerophosphoethanolamine-*N*-arachidonyl (NAPE 1; Fig. 6a) and alk-1-stearyl-2-arachidonyl-*sn*-glycerophosphoethanolamine-*N*-arachidonyl (NAPE 2; Fig. 6a). Identifications were based (1) on the occurrence of deprotonated molecules of appropriate mass (NAPE 1: mass-to-charge ratio (m/z) 1009; and NAPE 2, m/z 1039); and (2) on the chromatographic behaviour of these components, which was similar to that of synthetic NAPE (Fig. 6b; see Methods). NAPE 1 and NAPE 2 were synthesized by rat lung membranes in a Ca^{2+} -dependent manner. The membranes produced 1.5 ± 0.02 pmol of NAPE 1 and undetectable levels in NAPE 2 when incubated with EGTA (1 mM); and 4.4 ± 0.5 pmol of NAPE 1 and 3.1 ± 0.6 pmol of NAPE 2 when incubated with Ca^{2+} (3 mM) ($n = 4$; $P < 0.05$ between EGTA and Ca^{2+}) (Fig. 6b, insets). Guinea pig lung membranes also produced NAPE 1 and NAPE 2 in a Ca^{2+} -dependent

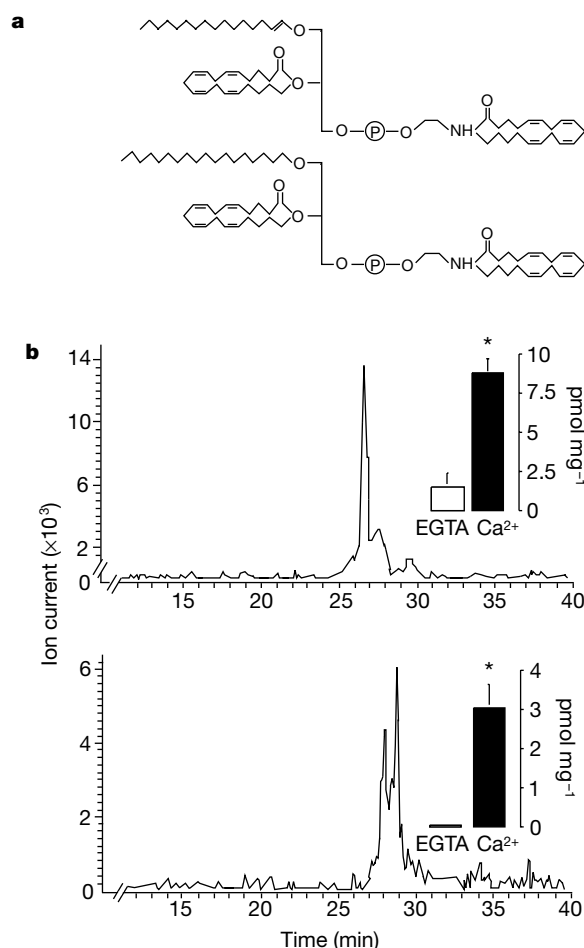


Figure 6 Ca^{2+} -dependent biosynthesis of anandamide precursors in rat lung. **a**, Structures of alk-1-palmito-2-arachidonyl-*sn*-glycero-phosphoethanolamine-*N*-arachidonyl (top panel, NAPE 1) and alk-1-stearoyl-2-arachidonyl-*sn*-glycero-phosphoethanolamine-*N*-arachidonyl (bottom panel, NAPE 2), two anandamide precursors. **b**, Representative HPLC/MS tracings for selected ions characteristic of NAPE 1 (top panel, $m/z = 1009$, deprotonated molecule, $[\text{M} - \text{H}]^-$) and NAPE 2 (bottom panel, $m/z = 1039$, $[\text{M} - \text{H}]^-$). The insets show that biosynthesis of NAPE 1 (top panel) and NAPE 2 (bottom panel) was significantly stimulated by Ca^{2+} (3 mM) (mean $\pm$ s.e.m.; asterisk, $P < 0.05$, $n = 4$).

manner (data not shown). The presence in rodent lungs of a Ca^{2+} -activated mechanism for the biosynthesis of anandamide and its phospholipid precursor supports a role for this endogenous cannabinoid in airway modulation.

Our results indicate that activation of CB1 receptors by locally released anandamide may participate in the control of bronchial contractility. How anandamide exerts such control may depend, however, on the state of the bronchial muscle. When the muscle is contracted, as during capsaicin-evoked bronchospasm, anandamide may counteract this contraction, possibly by inhibiting the prejunctional release of excitatory neurotransmitters and neuropeptides. In contrast, when the smooth muscle is relaxed, as seen after removal of the constricting influence of the vagus nerve, anandamide may cause bronchoconstriction. These state-dependent effects may account for the variable bronchomotor actions of acutely administered cannabis-like agents in humans, and might contribute to the long-term toxicity associated with cannabis smoking²⁴. Furthermore, the existence of an intrinsic cannabis-mediated control of airway function may open perspectives for the development of new antitussive and anti-asthmatic agents. For example, drugs targeting the mechanisms of anandamide

inactivation²⁵ may exert more selective pharmacological effects on bronchial responsiveness than those elicited by direct-acting CB1 agonists. □

Methods

Chemicals

Anandamide and [³H]anandamide were synthesized following standard procedures; 1-palmitoyl-2-oleyl-*sn*-glycero-phosphoethanolamine-*N*-arachidonyl and 1,2-dioleoyl-*sn*-glycero-phosphoethanolamine-*N*-oleyl were purchased from avanti Polar Lipids (Alabaster, AL); SR144528 (*N*-[1*S*]-endo-1,3,3-trimethyl bicyclo [2.2.1] heptan-2-yl]-5-(4-chloro-3-methylphenyl)-1-(4-methylbenzyl)-pyrazole-3-carboxamide) was a generous gift from Sanofi Recherche (Montpellier, France); SR141716A ([*N*-(piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1*H*-pyrazole-3-carboxamide-HCl] was provided by RBI (Natick, MA) as part of the Chemical Synthesis Program of the NIMH (N01MH30003); all other drugs were from Tocris (Ballwin, MO) or Sigma (Saint Louis, MO). The drugs were dissolved in dimethylsulphoxide (DMSO), and administered in physiological saline containing 10% DMSO.

Biological assays

For bronchospasm, we anaesthetized Dunkin-Hartley guinea-pigs (Charles-River, weighing 200–400 g) or Wistar rats (Charles-River, weighing 200–300 g) with pentobarbital (40 mg per kg intraperitoneal (i.p.)) and fentanyl (25 μg per kg intramuscular), catheterized the trachea and carotid artery to measure airway obstruction and systemic blood pressure, and catheterized the jugular vein to administer drugs. Pancuronium bromide (4 mg per kg i.v.) was administered to prevent spontaneous breathing. The animals were ventilated with room air using a rodent ventilator (U. Basile, Comerio, Italy) run at 60 strokes min^{-1} ; stroke volume was 3–7 ml. Airway resistance was measured by using a differential pressure transducer (U. Basile) connected by the side-arm of the tracheal catheter to a bronchospasm transducer. Bronchospasm was expressed as a per cent of the maximal response, which was determined by clamping the tracheal catheter before and after each experiment. Drugs were dissolved in saline containing 10% DMSO, and injected through the jugular vein; responses were evaluated at their peak. Arterial blood pressure was measured continuously with a pressure transducer connected to a recorder (U. Basile). To abolish vagal influences on bronchial musculature, in some experiments we bilaterally transected the vagus nerves and administered atropine sulphate (2 mg per kg i.v.). Anandamide was administered 15 min before capsaicin; the antagonists were given 15 min before anandamide.

For coughing, we individually exposed conscious guinea-pigs to aerosolized capsaicin (0.3 mM) for 4 min while recording coughs by using a microphone placed in the exposure chamber²⁶. Each animal was treated only once with capsaicin. Anandamide was administered either systemically (i.p. 30 min before capsaicin) or locally (by aerosol, 10 mg per ml, 15 min before capsaicin). Aerosols were prepared with an Air Lister Basic apparatus (Hatü, Italy), regulated at an emission flow rate of 6 l min^{-1} . For isolated lung strips, guinea-pig parenchymal strips were prepared essentially as described²⁷. The strips were placed in 10-ml organ baths containing Krebs' buffer (in mM: NaCl, 118; KCl, 4; K_2HPO_4 , 1.2; MgSO_4 , 1.2; CaCl_2 , 2.5; NaHCO_3 , 25.0; glucose, 11.2; supplemented with 7 μM atropine sulphate and 15 μM indomethacin) at 37 °C and aerated with an oxygen/carbon dioxide mixture (95/5%). Muscle contractions were recorded with an isometric force transducer (U. Basile) and are expressed in dyne per mg of fresh tissue.

Electron microscopy

The lungs were removed from three rats perfused with a phosphate-buffered (0.1 M) fixative containing 4% paraformaldehyde, 0.2% picric acid and 0.05% glutaraldehyde, and were further fixed for 24 h. We carried out immunohistochemical analyses as described¹⁹. Rabbit C-terminal anti-CB1 and anti-NPY antibodies were used at 1:5,000 and 1:20,000 dilution, respectively. The specificity of the NPY antibody was reported previously²⁸; the specificity of the CB1 antibody has been verified in several ways, including staining of CB1 knockout mice, and will be detailed elsewhere (N. Hájós, I. K., C. Ledent, K. M. and T. F. E., manuscript in preparation).

Membrane preparation and lipid extraction

Lung particulate fractions were prepared as described²⁹. We carried out incubations for 1 h at 37 °C in Tris buffer (50 mM, pH 7.4) containing CaCl_2 (3 mM) or EGTA (1 mM) and 2 mg ml^{-1} of membrane protein. Reactions were stopped by adding cold methanol, and the lipids were extracted with chloroform. Before HPLC/MS analysis, we fractionated anandamide and NAPE by silica gel column chromatography²³.

HPLC/MS

Anandamide was identified and quantified by reversed-phase HPLC coupled to positive ionization electrospray MS, by using an isotope-dilution method as described³⁰. We purified NAPE species by reversed-phase HPLC on a C_{18} Bondapak column (300 $\times$ 3.9 mm internal diameter 5 μm) (Waters) maintained at 20 °C and interfaced with an Agilent HP1100 model mass spectrometer. HPLC conditions consisted of a linear gradient of methanol in water (from 75 to 100% methanol in 30 min) with a flow rate of 1 ml min^{-1} . Under these conditions, different NAPE species were eluted from the column as a group of peaks at retention times comprising 27–29 min. MS analyses were performed with the electrospray ion source set in the negative ionization mode; the V_{cap} set at 5 kV; and the fragmentor voltage set at 200 V. Nitrogen was used as a drying gas at a flow rate of

12 l min⁻¹. The drying gas temperature was set at 350 °C, and the nebulizer pressure at 30 PSI. For quantitative purposes, we extracted diagnostic ions (deprotonated molecules, [M - H]⁻) from full scan data and quantified them by comparison with an external standard (1-palmityl-2-oleyl-*sn*-glycero-phosphoethanolamine-*N*-arachidonyl, Avanti Polar Lipids). NAPE 2 was eluted from the column as a doublet, and the areas under both peaks were combined for quantification.

Data analysis

Results are expressed as means ± s.e.m. The significance of differences among groups was evaluated using Student's *t*-test or analysis of variance followed by Dunnett's test.

Received 8 June; accepted 7 August 2000.

1. Vachon, L., Fitzgerald, M. X., Soliday, N. H., Gould, I. A. & Gaensler, E. A. Single-dose effects of marijuana smoke. Bronchial dynamics and respiratory-center sensitivity in normal subjects. *New Engl. J. Med.* **288**, 985–989 (1973).
2. Tashkin, D. P., Shapiro, B. J., Lee, Y. E. & Harper, C. E. Effects of smoked marijuana in experimentally induced asthma. *Am. Rev. Respir. Dis.* **112**, 377–386 (1975).
3. Tashkin, D. P. *et al.* Bronchial effects of aerosolized delta 9-tetrahydrocannabinol in healthy and asthmatic subjects. *Am. Rev. Respir. Dis.* **115**, 57–65 (1977).
4. Iversen, L. L. *The Science of Marijuana* (Oxford Univ. Press, Oxford, 2000).
5. Abboud, R. T. & Sanders, H. D. Effect of oral administration of delta-9-tetrahydrocannabinol on airway mechanics in normal and asthmatic subjects. *Chest* **70**, 480–485 (1976).
6. Devane, W. *et al.* Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* **258**, 1946–1949 (1992).
7. Di Marzo, V. *et al.* Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* **372**, 686–691 (1994).
8. Szallasi, A. & Blumberg, P. M. Vanilloid (Capsaicin) receptors and mechanisms. *Pharmacol. Rev.* **51**, 159–212 (1999).
9. Barnes, P. J. Modulation of neurotransmission in airways. *Physiol. Rev.* **72**, 699–729 (1992).
10. Rinaldi-Carmona, M. *et al.* SR141716A, a potent and selective antagonist of the brain cannabinoid receptor. *FEBS Lett.* **350**, 240–244 (1994).
11. Rinaldi-Carmona, M. *et al.* SR144528, the first potent and selective antagonist of the CB2 cannabinoid receptor. *J. Pharmacol. Exp. Ther.* **284**, 644–650 (1998).
12. Calignano, A., La Rana, G., Giuffrida, A. & Piomelli, D. Control of pain initiation by endogenous cannabinoids. *Nature* **394**, 277–281 (1998).
13. Karlsson, J. -A. & Fuller, R. W. Pharmacological regulation of the cough reflex—from experimental models to antitussive effects in man. *Pulmonary Pharmacol. Ther.* **12**, 215–228 (1999).
14. Widdicombe, J. G. Afferent receptors in the airways and cough. *Respir. Physiol.* **114**, 5–15 (1998).
15. Zygmunt, P. M., Julius, L., Di Marzo, V. & Hogestatt, E. D. Anandamide—the other side of the coin. *Trends Pharmacol. Sci.* **21**, 43–44 (2000).
16. Ishac, E. J. *et al.* Inhibition of exocytotic noradrenaline release by presynaptic cannabinoid CB1 receptors on peripheral sympathetic nerves. *Br. J. Pharmacol.* **118**, 2023–2028 (1996).
17. Coutts, A. A., Pertwee, R. G., Fernando, S. R. & Nash, J. E. Inhibition by cannabinoid receptor agonists of acetylcholine release from guinea-pig myenteric plexus. *Br. J. Pharmacol.* **121**, 1557–1566 (1997).
18. Richardson, J. D., Kilo, S. & Hargreaves, K. M. Cannabinoids reduce hyperalgesia and inflammation via interaction with peripheral CB1 receptors. *Pain* **75**, 111–119 (1998).
19. Katona, I. *et al.* Presynaptically located CB1 cannabinoid receptors regulate GABA release from axon terminals of specific hippocampal interneurons. *J. Neurosci.* **19**, 4544–4558 (1999).
20. Rice, W., Shannon, J. M., Burton, F. & Fiedelley, D. Expression of a brain-type cannabinoid receptor (CB1) in alveolar type II cells in the lung: regulation by hydrocortisone. *Eur. J. Pharmacol.* **327**, 227–232 (1997).
21. Landsman, R. S., Burke, T. H., Consroe, P., Roeske, W. R. & Yamamura, H. I. SR141716A is an inverse agonist at the human cannabinoid CB1 receptor. *Eur. J. Pharmacol.* **334**, R1–R2 (1997).
22. Sugiura, T. *et al.* Transacylase-mediated and phosphodiesterase-mediated synthesis of N-arachidonoylethanolamine, an endogenous cannabinoid-receptor ligand, in rat brain microsomes. *Eur. J. Biochem.* **240**, 53–62 (1996).
23. Cadas, H., di Tomaso, E. & Piomelli, D. Occurrence and biosynthesis of endogenous cannabinoid precursor, N-arachidonoyl phosphatidylethanolamine, in rat brain. *J. Neurosci.* **17**, 1226–1242 (1997).
24. Van Hoozen, B. E. & Cross, C. E. Marijuana. Respiratory tract effects. *Clin. Rev. Allergy Immunol.* **15**, 243–269 (1997).
25. Piomelli, D., Giuffrida, A., Calignano, A. & Rodriguez de Fonseca, F. The endocannabinoid system as a target for therapeutic drugs. *Trends Pharmacol. Sci.* **21**, 218–224 (2000).
26. Bolster, D. C., Del Prado, M., O'Reilly, S., Mingo, G. & Hey, J. A. Evan's blue dye blocks capsaicin-induced cough and bronchospasm in the guinea pig. *Eur. J. Pharmacol.* **276**, R1–R3 (1995).
27. Samhoun, M. N. & Piper, P. J. The combined use of isolated strips of guinea-pig lung parenchyma and ileum as a sensitive and selective bioassay for leukotriene B₄. *Prostaglandins* **27**, 711–724 (1984).
28. Csiffary, A., Gorcs, T. J. & Palkovits, M. Neuropeptide Y innervation of ACTH-immunoreactive neurons in the arcuate nucleus of rats: a correlated light and electron microscopic double immunolabeling study. *Brain Res.* **506**, 215–222 (1990).
29. Désarnaud, F., Cadas, H. & Piomelli, D. Anandamide amidohydrolase activity in rat brain microsomes: Identification and partial characterization. *J. Biol. Chem.* **270**, 6030–6035 (1995).
30. Giuffrida, A., Rodriguez de Fonseca, F. & Piomelli, D. Quantification of bioactive acylethanolamides in rat plasma by electrospray mass spectrometry. *Anal. Biochem.* **280**, 87–93 (2000).

Acknowledgements

We thank S. Chen, M. Elefante, G. Fabozzi, G. Goda, E. Oswald and R. Russo for excellent experimental assistance; T. Dinh, H. Kim and F. Nava for reading the manuscript critically; and B. Vigh for discussion. This research was supported from the National Institute of Drug Abuse (D.P. and K.M.), by the Howard Hughes Medical Institute and OTKA (T.F.), and by MURST (A.C.).

Correspondence and requests for materials should be addressed to D.P. (e-mail: piomelli@uci.edu).

Identification of genes that modify ataxin-1-induced neurodegeneration

Pedro Fernandez-Funez^{*†}, Maria Laura Nino-Rosales^{*†}, Beatrice de Gouyon^{*}, Wei-Chi She^{*‡}, James M. Luchak^{*}, Pedro Martinez^{*}, Enrique Turiegano[§], Jonathan Benito[§], Maria Capovilla^{*}, Pamela J. Skinner^{||}, Alanna McCall^{*}, Inmaculada Canal[§], Harry T. Orr^{||}, Huda Y. Zoghbi^{*‡§#} & Juan Botas^{*‡}

^{*} Department of Molecular and Human Genetics, [†] Department of Pediatrics, [‡] Program in Developmental Biology, and [#] Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

[§] Departamento de Biología, Universidad Autónoma de Madrid, 28049 Madrid, Spain

^{||} Institute of Human Genetics, University of Minnesota, Minneapolis, Minnesota 55455, USA

[†] These authors contributed equally to this work

A growing number of human neurodegenerative diseases result from the expansion of a glutamine repeat in the protein that causes the disease¹. Spinocerebellar ataxia type 1 (SCA1) is one such disease—caused by expansion of a polyglutamine tract in the protein ataxin-1. To elucidate the genetic pathways and molecular mechanisms underlying neuronal degeneration in this group of diseases, we have created a model system for SCA1 by expressing the full-length human SCA1 gene in *Drosophila*. Here we show that high levels of wild-type ataxin-1 can cause degenerative phenotypes similar to those caused by the expanded protein. We conducted genetic screens to identify genes that modify SCA1-induced neurodegeneration. Several modifiers highlight the role of protein folding and protein clearance in the development of SCA1. Furthermore, new mechanisms of polyglutamine pathogenesis were revealed by the discovery of modifiers that are involved in RNA processing, transcriptional regulation and cellular detoxification. These findings may be relevant to the treatment of polyglutamine diseases and, perhaps, to other neurodegenerative diseases, such as Alzheimer's and Parkinson's disease.

Drosophila provides a flexible and powerful model to study neurodegenerative diseases. Using the GAL4/UAS system², we can control the level of transgene expression and direct expression to different cell types or even to specific neurons. Large-scale genetic screens allow us to identify the genes and pathways involved in pathogenesis. Because most of the genetic pathways involved in normal development and disease conditions are conserved between *Drosophila* and mammals, mechanisms of neuronal degeneration in *Drosophila* may prove relevant to neurodegeneration in humans. *Drosophila* models using polyglutamine or truncated polypeptides of ataxin-3 and huntingtin, for example, show the characteristic progressive neural degeneration and pathology^{3–6}, and overproduction of the Hsp70 and Hsp40 molecular chaperones suppresses polyglutamine-induced neurotoxicity^{3,5}.

Because polyglutamine peptides tend to be more toxic than the full-length proteins and do not elicit the cell-type-specific neurodegeneration characteristic of these diseases⁷, their activities might be distinct from those of the full-length proteins. Therefore, we generated transgenic flies expressing full-length ataxin-1 using the GAL4/UAS system. We cloned two human SCA1 complementary DNAs, differing in the size of their polyglutamine repeat tracts, into the *Drosophila* transformation vector pUAST. These constructs encoded ataxin-1 30Q (a wild-type human isoform) and ataxin-1 82Q (an expanded isoform) (Fig. 1). Six 82Q and four 30Q lines were generated.

We directed SCA1 expression to the eye retina using the *gmr*-GAL4 driver⁸ and estimated ataxin-1 levels in all transgenic lines by

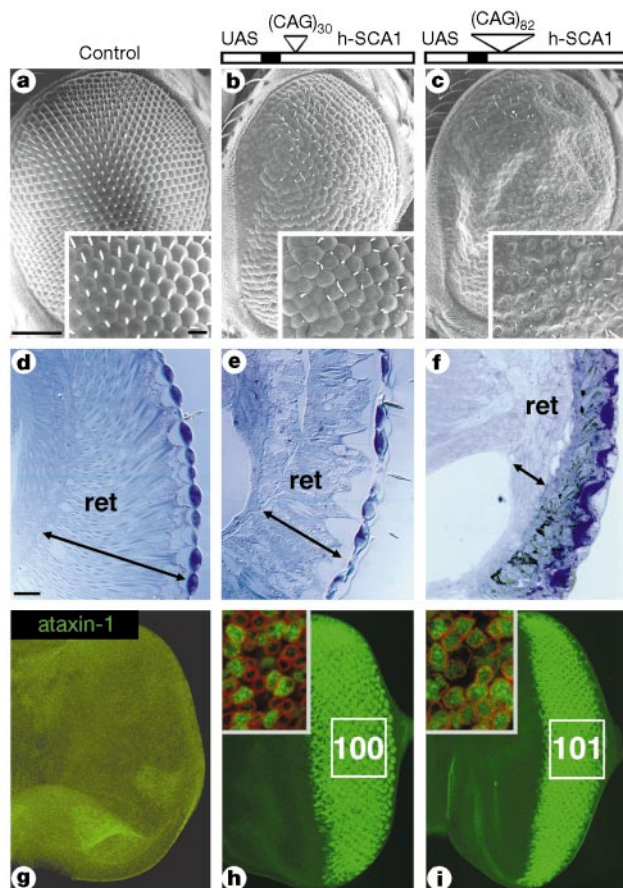


Figure 1 Strong (ataxin-1 82Q) and weak (ataxin-1 30Q) eye phenotypes produced by SCA1 overexpression. **a–c**, SEM eye images of *gmr–GAL4/+* control (**a**), *gmr–GAL4/+; UAS: SCA1 30Q[F6]/+* (**b**) and *gmr–GAL4/+; UAS: SCA1 82Q[M6]/+* (**c**) transgenic flies raised at 23 °C. Insets, magnification of ommatidia field. **d–f**, Sections through the eye retinas of *gmr–GAL4/+* control (**d**), *gmr–GAL4/+; SCA1 30Q[F6]/+* (**e**) and *gmr–GAL4/+; SCA1 82Q[M6]/+* (**f**) transgenic flies at identical magnification. Degeneration of the retina (arrow) is severe in 82Q[M6] flies, and milder in 30Q[F6] flies. **g–i**, Third instar eye imaginal discs stained with anti-ataxin-1 antiserum. No protein is detected in controls. Ataxin-1 30Q and ataxin-1 82Q are expressed at similar levels (numbers indicate the relative amounts of immunofluorescence in each line). Insets, magnification showing ataxin-1 nuclear inclusions (green) and the nuclear membrane (red) revealed by anti-laminin antibody. Scale bars: whole eye, 100 μ m; eye inset and retinal section, 10 μ m.

immunofluorescence (see Methods). The severity of the phenotype correlated well with expression levels in both the 82Q and the 30Q transgenic lines, although, as expected, the 82Q lines showed a much stronger phenotype. The most severe eye phenotypes of the 82Q and 30Q lines are shown in Fig. 1a–f. Note that these two lines have roughly similar expression levels (Fig. 1h, i).

Increasing the expression of ataxin-1 30Q by increasing the culture temperature leads to a more pronounced phenotype (Fig. 2a–d). Similar results were observed when the dosage of the transgene was doubled (data not shown). The finding that relatively high expression of wild-type human SCA1 transgenes can lead to phenotypes similar to those produced by expanded alleles was unexpected. Previous analysis of transgenic mice expressing the wild-type human SCA1 30Q allele revealed no obvious pathology⁹, but the expression level of the SCA1 30Q AO2 transgene was roughly one-quarter of that of the SCA1 82Q BO5 transgene, as determined by northern analysis¹⁰. To determine whether SCA1 30Q could also cause neurodegeneration in mice, we examined the cerebellar cortex of homozygous SCA1 30Q AO2 mice. The dendritic arborization of

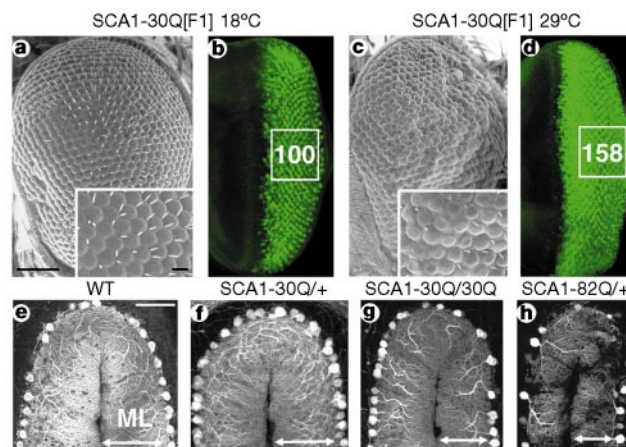


Figure 2 Ataxin-1 82Q and high levels of ataxin-1 30Q cause similar phenotypes, in both *Drosophila* and mice. **a, c**, SEM eye images from *UAS: SCA1 30Q[F1]/+; gmr–GAL4/+* flies raised at 18 °C (**a**) and 29 °C (**c**). **b, d**, Third instar eye imaginal discs stained with anti-ataxin-1 antiserum. Ataxin-1 expression at 29 °C is roughly 158% of that at 18 °C, as determined by immunofluorescence. Increased expression aggravates the phenotype. **e–h**, Mouse cerebellar sections stained with antisera against the Purkinje-cell-specific protein calbindin at identical magnifications: **e**, 51-week-old wild type; **f**, 52-week-old heterozygous SCA1 30Q[AO2]; **g**, 59-week-old homozygous SCA1 30Q[AO2]; **h**, 50-week-old heterozygous SCA1 82Q[BO5]. Scale bars: whole eye, 100 μ m; eye inset, 10 μ m; cerebellar sections, 100 μ m.

the Purkinje cell layer is clearly abnormal at 59 weeks in these mice, and resembles the degeneration seen at 12–15 weeks in the SCA1 82Q BO5 transgenic line (Fig. 2e–h). The only difference is the length of time required for pathology to appear. Thus, even wild-type unexpanded human ataxin-1 can produce a neurodegenerative phenotype if expressed at sufficiently high levels.

To determine whether SCA1 neurodegeneration in flies is progressive, we used the Δ^{VNC} GAL4 driver, which directs expression to a small number of well-defined interneurons in the ventral nerve cord of the adult central nervous system (CNS). We assessed the integrity of the ventral cord interneurons at days 1, 25 and 45 of adult life in wild-type and transgenic flies expressing ataxin-1 30Q or 82Q. Both the cell body and axonal projections of these interneurons can be easily visualized with the τ -GFP reporter gene (Fig. 3a, b). We found a progressive elimination of the cell bodies and axonal projections with age on ataxin-1 82Q expression (Fig. 3). Among the ataxin-1 30Q lines, we observed only mild neuronal loss in the line with highest expression (data not shown). Expression of ataxin-1 in fly neurons therefore causes progressive degeneration as in patients with SCA1 and transgenic mice.

We observed nuclear inclusions in a variety of cell types expressing 30Q or 82Q ataxin-1, including eye photoreceptor cells (see insets, Fig. 1h, i), neurons of the CNS (Fig. 3c, d) and cells of the mature salivary glands. As in mammalian cells, ataxin-1 nuclear inclusions in *Drosophila* are positive for the Hsp70 chaperone, ubiquitin and the proteasome (see Supplementary Information). As ataxin-1 is degraded by the ubiquitin–proteasome pathway¹¹ and has been shown to redistribute certain chaperones in patient tissue, we expected that reducing chaperone or proteasome activity would aggravate the neurodegenerative phenotype. We used existing mutations in *Drosophila* genes encoding molecular chaperones or components of the proteasome and crossed them with flies expressing a SCA1 82Q transgene. Control heterozygous flies carrying these mutations alone show a wild-type eye phenotype (Fig. 4f; and data not shown), but SCA1 82Q[F7] transgenic flies heterozygous for any of these mutations show a more severe eye phenotype than flies carrying only the SCA1 82Q transgene (Fig. 4a–e).

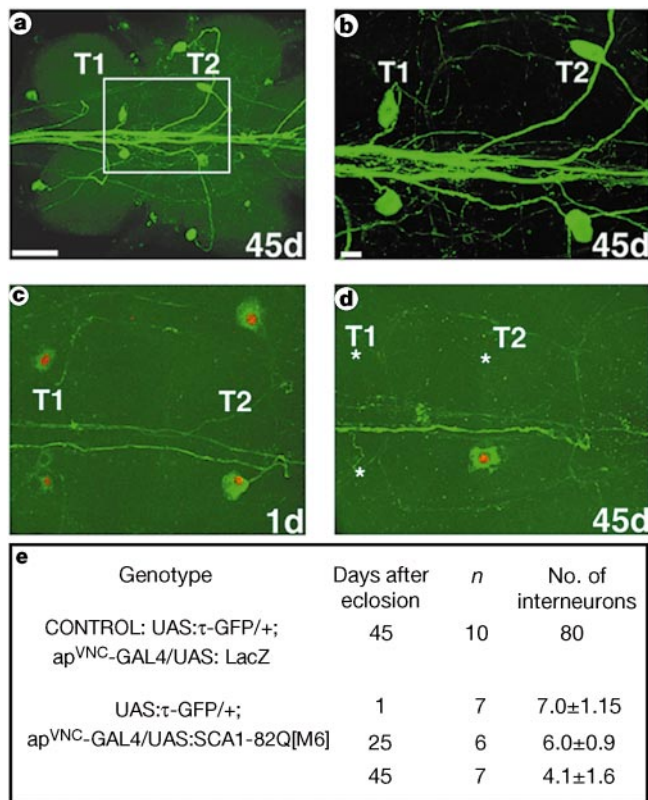


Figure 3 Expression of ataxin-1 82Q in *Drosophila* interneurons causes progressive degeneration. All panels show *apterous*-expressing interneurons (two per hemisegment) in the first (T1) and second (T2) thoracic segments of the adult CNS, visualized with the τ-GFP reporter gene driven by the ap^{VNC}GAL4 driver. **a**, CNS of 45-day-old control ap^{VNC}GAL4/+; UAS: τ-GFP/+; UAS: lacZ/+ fly shows robust axonal projections of interneurons that fasciculate with axons from interneurons in other segments. **b**, Magnification of **a** showing two T1 and two T2 ventral interneurons and their axonal projections. **c**, **d**, magnification images of 1-day-old (**c**) and 45-day-old (**d**) ap^{VNC}GAL4/+; UAS: τ-GFP/+; UAS: SCA1 82Q[M6]/+ flies. Ataxin-1 82Q accumulates in nuclear inclusions (red label). Most cell bodies are present at day 1 although label in axonal projections is already weak (these interneurons form several days earlier during metamorphosis). By day 45, fewer cell bodies are visible. **e**, Numbers of interneurons at different ages are shown for control and 82Q genotypes. In controls, the interneurons were scored by expression of GFP or LacZ. SCA1-expressing neurons were scored by positive staining with anti-SCA1 antibody. Comparing the number of interneurons: $P < 0.005$ between days 1 and 45; $P < 0.025$ between days 25 and 45.

To identify other genes capable of modifying SCA1-induced neurodegeneration, we carried out two genetic screens. First, we crossed a collection of 1,500 lethal P-element insertions with a UAS: SCA1[82Q]F7;gmr-GAL4/CyO strain. In these flies, a moderate level of ataxin-1 82Q accumulates in the retina, causing an intermediate eye phenotype at 25 °C. We then carried out a second genetic screen to identify genes that when overexpressed suppress or enhance SCA1-induced neurodegeneration by crossing the UAS:SCA1[82Q]F7;gmr-GAL4/CyO strain transgenic flies with a collection of 3,000 EP insertions¹². Putative modifiers detected in the initial screens were tested again in different conditions (see Methods) to verify their phenotype, determine their specificity (that is, modifiers altering the eye phenotype of SCA1 transgenic but not control sibling flies) and test whether they alter GAL4-ataxin-1 expression levels. We further studied only those modifiers that passed all these tests.

The P-element F₁ screen produced 7 suppressors and 20 enhancers of the SCA1 eye phenotype; that is, genes suppressing or

enhancing SCA1-induced neurodegeneration when their activity is reduced up to 50%. The EP-element F₁ screen produced 10 suppressors and 23 enhancers of the SCA1 eye phenotype. In principle, the eye phenotype modifications in the EP element screen could be caused by overexpression of a nearby transcription unit, but lack of function caused by insertional mutagenesis underlies some modifiers (Table 1). We report those SCA1 modifiers caused by insertional mutagenesis whose identity could be verified by using other available alleles. For some uncharacterized genes, we generated new alleles by ethylmethane sulphonate (EMS) mutagenesis or imprecise excision of the EP element. We also report some modifiers (P and EP) for which independent alleles are unavailable for verification; we chose only those that fall into classes defined by other modifiers and in which sequence analysis reveals only one gene as a candidate. In the case of the remaining modifiers, no independent alleles are available, and the P or EP insertions do not unambiguously identify the corresponding gene; thus, further analyses will be required.

Some of the identified modifiers are genes in the protein folding/heat-shock response or ubiquitin-proteolytic pathways (Table 1 and Fig. 4), and are known to be involved in polyglutamine-induced neurodegeneration. P1666 (Fig. 4g) is an insertion in a gene encoding ubiquitin (*Ubi63E*). We generated two EMS alleles of P1666, and these mutations also behave as enhancers of the eye phenotype (not shown). A mutation (*Uch-L3*^{2B8}) in a ubiquitin carboxy-terminal hydrolase, a protein involved in the recycling of ubiquitin, also enhances the SCA1 eye phenotype (data not shown). P1779 (Fig. 4h) and EP674 (data not shown) are mutations in *UbcD1/effete*, a gene that encodes a ubiquitin conjugase homologous to human UbcE2D2 (93% identity) and yeast Ubc4/5 (81% identity). We used another allele of *UbcD1/effete* to verify the interaction (Table 1). EP1303 (Fig. 4i) identifies a different ubiquitin conjugase, previously uncharacterized in flies, that is similar to human Ubc2EH (72% identity) and *Saccharomyces cerevisiae* Ubc8 (57% identity). P292, which is associated with a mutation in a heat-shock response factor known as *hsr-ω*, is an enhancer (Fig. 4j). Another existing allele of *hsr-ω* also enhances SCA1 neurodegeneration (Table 1). EP411 is a suppressor (compare Fig. 4k and l with m and n) associated with overexpression of a *Drosophila* DNAJ-1 gene (*dDNAJ-1* 64EF). This gene encodes a protein homologous to the human chaperone HSP40/HDJ-1 (50% identity over 117 amino acids). Both the *Drosophila* and human genes suppress polyglutamine toxicity^{5,13}. Notably, we also found the nuclear inclusions are altered in *dDNAJ-1* 64EF overexpressing flies (verified in two independent lines). Nuclear inclusions in these flies are more compact (Fig. 4p) and occupy a smaller portion of the nucleus than the typical ataxin-1 82Q control nuclear inclusions (Fig. 4o). Overall, they resemble the nuclear inclusions characteristic of ataxin-1 30Q (see Supplementary Information).

We recovered three suppressors that identify genes and pathways not previously known to be involved in polyglutamine-induced neurodegeneration (Table 1). EP2231 (Fig. 5b, f) causes the overproduction of a glutathione-S-transferase (GST) gene (*Gst55F*) that is most similar to human GSTs of the θ-class. GSTs are enzymes that have a principal role in cellular detoxification, and use reduced glutathione in conjugation and reduction reactions with a variety of toxins, including products of chemical and oxidative stress¹⁴. We generated imprecise excisions of EP2231 that enhance the SCA1 phenotype (Fig. 5j). We also obtained two lack-of-function mutations (P1480 and P874) in *Gst2*, a different *Gst* gene mapping to chromosomal location 53F and most similar to the human GST σ-class. These mutations also enhance the SCA1 eye phenotype (see Supplementary Information). These findings implicate oxidative and/or chemical stress in polyglutamine neurodegeneration. EP2417 (Fig. 5c, g) is associated with overexpression of an uncharacterized *Drosophila* gene that we named *nucleoporin-44A* (*nup-44A*). It encodes a protein homologous to the *S. cerevisiae* nuclear

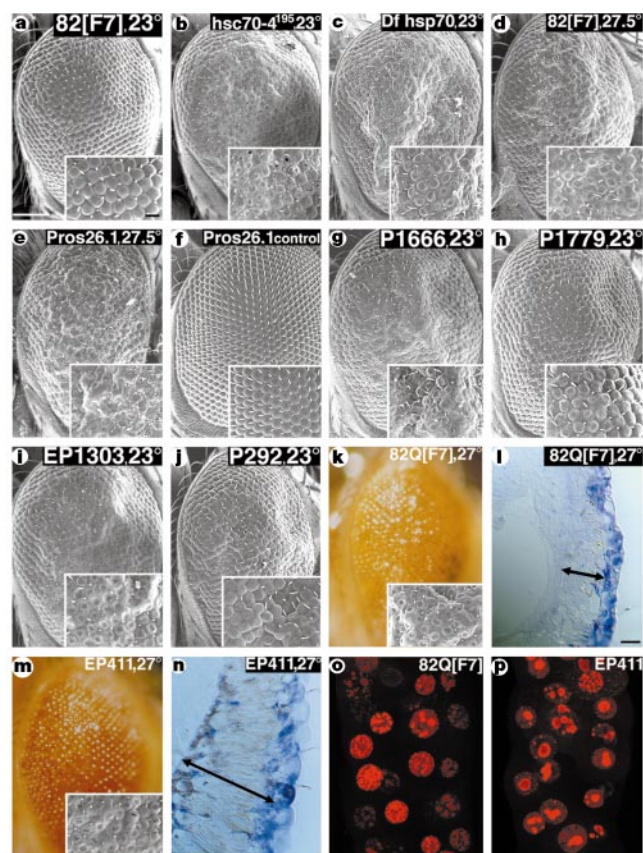


Figure 4 Modifiers of ataxin-1 82Q neurodegeneration in the protein folding/heat-shock response and ubiquitin proteolytic pathways. **a**, Eye phenotype of 82Q control fly (*UAS: SCA1 82Q[F7]/+*; *gmr-GAL4/+*) raised at 23°C. **b, c**, Eye phenotypes of flies heterozygous for *hsc70-4*¹⁹⁵ (**b**), and *DF(3R)karD1*, a small deletion removing a cluster of *hsp70* genes (*hsp70Ab*, *hsp70Ba*, *hsp70Bb* and *hsp70Bc*) (**c**). **d**, 82Q control fly raised at 27.5°C. **e**, 82Q fly heterozygous for *Pros26-DTS*. **f**, *Pros26-DTS/+* control. **g–j**, 82Q flies heterozygous for P1666, a mutation in *Ubiquitin 63E* (**g**), P1779, a mutation in the ubiquitin conjugase *UbcD1* (**h**), EP1303, an insertion in the ubiquitin conjugase *dUbcE2H* (**i**), P292, a mutation in the heat-shock response factor *hsc- ω* (**j**). **k, l**, Fresh eye image, SEM image (inset) and retinal section show rough eye surface, disorganized ommatidia and retinal thickness in 82Q control flies raised at 27°C. **m, n**, Parallel images of 82Q fly bearing EP411 (overexpressing *DNAJ-1*) reveal less disrupted ommatidia and improved retinal thickness. **o**, Salivary glands of 82Q control fly stained with anti-ataxin-1 antiserum to reveal the nuclear inclusions. **p**, 82Q fly overexpressing *DNAJ-1 64F* shows more compact and less invasive nuclear inclusions. Scale bars: whole eye, 100 μ m; eye inset 10 μ m; retinal sections, 10 μ m.

pore protein SEH1 (34% identity over 185 amino acids). The nuclear pore complex comprises many proteins¹⁵; overexpression of one component may impair nuclear pore complex formation and thus impede ataxin-1 import into the nucleus.

Other modifiers of SCA1 neurodegeneration are genes that encode proteins containing RNA-binding domains. For example, EP3623 (Fig. 5d, h) is a suppressor that overexpresses *mushroom body expressed (mub)*, which encodes a protein similar to vertebrate RNA-binding KH-domain proteins, suggesting that *mub* stabilizes specific messenger RNAs¹⁶. A *mub* lack-of-function allele does not modify the eye phenotype (Table 1). Notably, the other genes encoding proteins containing RNA-binding domains (Table 1, Fig. 5) are enhancers. These findings suggest that alteration of RNA processing is relevant to SCA1 pathogenesis. Consistent with this is the observation that ataxin-1 binds RNA *in vitro*, suggesting that ataxin-1 is an RNA binding protein itself (S. Yue and H.T.O., unpublished data). One of the modifiers in this class is a suppressor,

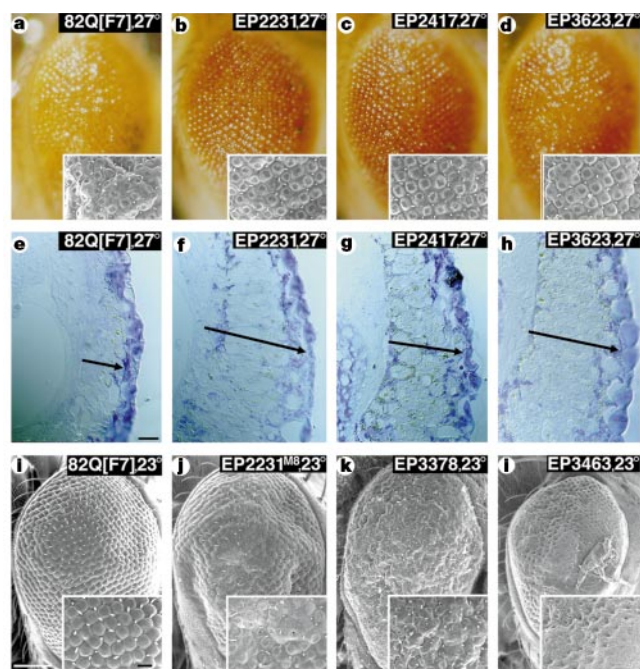


Figure 5 Suppressors and enhancers of ataxin-1 82Q neurodegeneration in new pathways. **a, e**, Fresh eye and SEM (inset) (**a**), and retinal section (**e**) showing the eye phenotype of control (*UAS: SCA1 82Q[F7]/+; gmr-GAL4/+*) flies; temperatures at which flies were raised are given. Arrow shows the retina. These phenotypes are partially suppressed in **b–d** and **f–h**, respectively, which show flies of the same genotype as controls with the following modifications: **b, f**, EP2231 (overexpressing *Gst55F*). **c, g**, EP2417 (overexpressing *nup44A*). **d, h**, EP3623 (overexpressing *mub*). Note changes in retinal thickness, illustrating suppression of phenotype. **i–k**, Mild eye phenotype of controls (*UAS: SCA1 82Q[F7]/+; gmr-GAL4/+*) (**i**) is aggravated in flies bearing EP2231^{MB}, an imprecise excision of EP2231 (**j**), EP3378, overexpressing *cpo* (**k**), and also carrying EP3463, a mutation in *tara* (**l**). Scale bars as in Fig. 4.

suggesting that it may be possible to slow SCA1 neurodegeneration by altering the activities of specific molecules involved in RNA processing.

A second group of enhancers are proteins that function as cofactors in transcriptional regulation. These include Sin3A, Rpd3 and dCtBP (see Table 1 and Fig. 5). These modifiers suggest that abnormal interactions between ataxin-1 and the transcriptional machinery may be another pathway for pathogenesis. In this context, alterations in gene expression occur very early during SCA1 pathogenesis¹⁷. Perturbation of the proteolytic machinery might alter levels of important transcription factors whose concentrations are regulated by proteolysis; alternatively, mutant ataxin-1 may interfere with nuclear domains important for transcriptional regulation¹⁷ or ataxin-1 may directly interact with certain components of the transcriptional machinery. Relatively short polyglutamine tracts are found in many transcription factors; thus, SCA1 and other polyglutamine disease proteins may interfere with specific transcriptional regulators (see ref. 18). This last possibility is supported by the finding that all modifiers in this class are transcriptional cofactors, but not other components of the transcriptional machinery (we directly tested components of polymerase II and the *Polycomb* group genes and found no effect). Also supporting this model is the recent observation that the amino terminus of huntingtin interacts with the nuclear receptor corepressor (N-CoR) in the yeast two-hybrid system¹⁹.

We investigated whether the SCA1 82Q modifiers described above also modify SCA1 30Q toxicity. Although the suppressors do not modify the disorganized ommatidial arrangement of SCA1

Table 1 Suppressors and enhancers of SCA1

Line	Modification*	Gene	Function	Accession no.	Insertion†	EP orientation/ overexpression‡	LOFA/M§
P1666	En	<i>Ubi63E</i>	Ubiquitin	M22428	–1,913	–	Yes (two EMS-induced alleles of P1666)/En
P1779	En	<i>UbcD1</i>	Ubiquitin conjugase	X62575	–432	–	Yes (<i>UbcD1</i> ⁵¹⁷⁸²)/En
EP674	En	<i>UbcD1</i>	Ubiquitin conjugase	X62575	–484	O/No	Yes (<i>UbcD1</i> ⁵¹⁷⁸²)/En
EP1303	En	<i>dUbc-E2H</i>	Ubiquitin conjugase	CG2257	–1,439	O/No	n.a.
P292	En	<i>hsc-ω</i>	Heat-shock response factor	U18307	–240	–	Yes (<i>hsc-ω</i> ⁵²⁴¹)/En
EP411	Su	<i>DnaJ1 64EF</i>	Chaperone	U34904	–600	S/Yes	n.a.
EP2231	Su	<i>Gst-θ1 55F</i>	Glutathione S-transferase	CG5164	–78	S/Yes	Yes (imprecise excision of EP2231)/En
EP2417	Su	<i>nup44A</i>	Nucleoporin	CG8722	–257	S/Yes	n.a.
EP3623	Su	<i>mub</i>	RNA binding	X99340	A	S/Yes	Yes (<i>mub</i> ⁰⁴⁹⁰³³)/none
EP3461	En	<i>pum</i>	RNA binding	L07943	B	S/Yes	Yes (<i>pum</i> ¹³)/none
EP3378	En	<i>cpo</i>	RNA binding	Z14311	–685	S/Yes	Yes (<i>cpo</i> ¹ and <i>cpo</i> ^{L1})/none
EP3725	En	<i>dY7521-B</i>	RNA binding	AF145594	–437	S/Yes	n.a.
EP866	En	<i>Sin3A</i>	Transcriptional cofactor	AJ007518	+4,601	O/No	Yes (<i>Sin3A</i> ^{d04} and <i>Sin3A</i> ⁰⁸²⁶⁹)/En
EP3672	En	<i>Rpd3</i>	Transcriptional cofactor	AF086715	–1,363	S/No	Yes (<i>Rpd3</i> ⁰⁴⁵⁵⁶)/En
P1590	En	<i>dCtBP</i>	Transcriptional cofactor	AJ224690	–7,353	–	Yes (<i>dCtBP</i> ^{8/De-10})/En
EP2300	En	<i>dSir2</i>	Transcriptional cofactor	AF068758	–427	S/Yes	Yes (<i>dSir2</i> ⁰⁵³²⁷)/None
P198	En	<i>Pap/Trap</i>	Transcriptional cofactor	AF226855	+800	–	Yes (<i>pap</i> ⁵³ and <i>pap</i> ^{EP1})/En
EP3463	En	<i>tara</i>	Transcriptional cofactor	AF227213	C; –553 + 16,335	S/D	Yes (five imprecise excisions of EP3463)/En

* En, enhancer; Su, suppressor.

† Insertion refers to insertion site relative to putative ATG at +1. A, inserted 22 base pairs into 5' non-coding exon, separated from ATG by ~28-kb intron; B, inserted in intron between exons 8 and 9; C, inserted in intron ~16.3-kb downstream of first ATG, but only ~553 with respect of second ATG; D, isoform 1A (starting at +16,335) is disrupted, whereas isoform 1B (starting at –553) is overexpressed.

‡ Orientation of P/EP relative to transcription unit. S, same; O, opposite.

§ LOFA, other loss-of-function alleles of the modifier gene; M, modification of SCA1 eye neurodegeneration caused by LOFA; none, no modification of SCA1 eye phenotype by loss of function allele (LOFA). n.a., not available.

30Q flies, they cause some recovery of the pigmentation, suggesting partial suppression of the phenotype in the non-neural pigment cells. The enhancers only weakly aggravated the 30Q phenotype, in contrast to their strong intensification of the 82Q phenotype (not shown).

The finding that the gain of ataxin-1 toxicity normally associated with polyglutamine expansion can also result from excess wild-type protein is intriguing. Overexpression of wild-type α -synuclein in transgenic mice leads to the formation of ubiquitin-positive inclusions and neural degeneration that resembles the Parkinsonian pathogenesis caused by mutant α -synuclein²⁰. One hypothesis is that the gain of ataxin-1 (and α -synuclein) toxicity results, in part, from protein misfolding. According to this model, ataxin-1 has more than one stable conformation, and a certain percentage of even the wild-type protein misfolds. This alternative stable conformation is pathogenic. Polyglutamine tract expansion increases the likelihood that the protein will adopt a misfolded, pathogenic conformation. Expanded ataxin-1 may also interact abnormally with other proteins and/or nucleic acids. The cell's proteolytic machinery is able to degrade some misfolded ataxin-1, but over a certain threshold—caused either by overabundant wild-type protein expression or by polyglutamine expansions—the misfolded protein is not cleared efficiently and induces neuronal dysfunction. This is in agreement with previous studies of ataxin-1 protein turnover¹¹. Modifiers that significantly affected the 82Q phenotype only mildly affected the 30Q phenotype. This argues that the polyglutamine tract has a pathogenic role of in the context of ataxin-1. Why, then, is ataxin-1 30Q not toxic in humans? The answer to this question may lie in our finding that the SCA1 phenotype is highly sensitive to expression levels. Ataxin-1 30Q in humans may never reach the critical level required for pathogenesis.

Our SCA1 *Drosophila* model provides insight into the multifaceted nature of SCA1 pathogenesis. The study of flies overexpressing SCA1 30Q and SCA1 82Q points to the role of protein levels in mediating pathogenesis and to the toxic effects of the expanded polyglutamine tract. The discovery of modifiers involved in GST-mediated cellular detoxification, transcriptional regulation and RNA processing reveals additional pathogenic mechanisms in spinocerebellar ataxia type 1. These modifiers may be relevant to future studies aimed at treating polyglutamine diseases, and, perhaps, other neurodegenerative diseases caused by proteins, such as Alzheimer's disease and Parkinson's disease. □

Methods

Drosophila genetics

Fly culture and crosses were carried out at 23 °C unless otherwise noted. The UAS: SCA1 30Q and UAS: SCA1 82Q transgenic flies were generated by cloning two human SCA1 complementary DNAs containing 30 and 82 CAG repeats, respectively⁹, in the pUAST transformation vector². These constructs were injected in a *y*¹*w*¹¹¹⁸ strain by standard methods. The EP strains were provided by C. Cater and G. Rubin; *hsc70*⁴¹⁹⁵ flies were provided by S. Artavanis-Tsakonas; *cpo*¹ and *cpo*^{L1} were provided by H. Bellen; *pap*⁵³ and *pap*^{EP1} were provided by D. Cribbs; The P strains and all other strains were provided by the Bloomington *Drosophila* Stock Center. For the modifier screens, females of the genotype *y*¹*w*¹¹¹⁸ UAS: SCA1 82Q[F7]; *gmr*–GAL4/CyO were crossed to males from the collection of lethal P insertions, EP insertions or other selected strains. Culture and crosses of the genetic screens were made at 25 °C. Strains producing a modification of the 82Q[F7] characteristic phenotype were kept and crossed again at 23 °C, 27 °C and 29 °C to generate different GAL4 expression levels. Because high expression of GAL4 itself at 29 °C causes an eye phenotype, we could control for possible modifiers altering Gal4 expression in siblings carrying *gmr*–GAL4 and the modifier, but not the UAS: SCA1 transgene. EP lines were tested for specificity by crossing females of selected EP lines with males bearing *y*¹*w*¹¹¹⁸UAS: SCA182Q[F7]; *gmr*–GAL4/CyO. As UAS: SCA182Q[F7] is on the X chromosome, F₁ males do not carry this transgene and serve as controls of EP specificity. To produce the genotypes shown in Fig. 3, we crossed UAS: τ -GFP; *ap*^{YNC}/CyO flies with UAS: SCA182Q[M6] flies or UAS: LacZ as controls. Adult flies were aged in standard medium after eclosion.

Histology and immunofluorescence

For scanning electron microscopy (SEM) images, whole flies were dehydrated in ethanol, critical-point dried and analysed with a JEOL JSM 6100 microscope. For sections of adult fly eyes, adult heads (0–1 days) were fixed for 30 min in 4% formaldehyde, washed, fixed for 3 h at 4 °C in 1% osmium tetroxide, dehydrated in ethanol, embedded in Epon for vertical semi-thin sections, and then stained with toluidine blue. SCA1-30Q and SCA1-82Q mice have been described⁹. Mouse cerebellar sections were prepared as described¹¹ and stained with an anti-calbindin monoclonal antibody (1: 1,000, CL300 Sigma).

Eye imaginal discs and salivary glands were dissected in 1× PBS, fixed for 20 min in 4% formaldehyde, washed with 1× PBS, 0.1% TritonX-100, and incubated with the primary antibody. Adult ventral ganglions were prepared by fixing the whole fly thorax for 3 h at 4 °C in 4% formaldehyde. After washing, the ventral ganglions were dissected and fixed again for 20 min at room temperature, and then stained as imaginal discs. Primary antibodies: rabbit anti-ataxin-1 (11NQ, diluted 1:750; ref. 11), mouse anti-laminin A (T47, 1:50; ref. 21), mouse anti-hsp70/hsc70 (SPA822, 1:100; StressGen), mouse anti-ubiquitin (Ubi-1, 1X; ZYMED), mouse anti-19S Regulator ATPase subunit 6b (Tbp7) (PW8175, 1:100; AFFINITI). Secondary fluorochrome-conjugated antibodies: Alexa⁴⁸⁸ (1:400, Molecular Probes) and Cy3 (1:500, Jackson). Confocal microscopy was carried out on Bio-Rad MRC-1024 confocal imaging system, and images were collected with Lasersharp 3.0 software (BioRad) and modified with Adobe Photoshop 4.0 (Adobe).

Comparison of ataxin-1 immunoreactivity

Eye imaginal discs to be compared, differing in genotype or temperature of culture, were manipulated in parallel. They were dissected in 1× PBS solution, fixed with 4% formaldehyde, incubated in rabbit anti-ataxin-1 antiserum 11NQ (dilute 1:750, ref. 11) for 14 h at 4 °C, washed in PBT (1× PBS, 0.1% TritonX-100), and incubated in goat anti-rabbit antibody conjugated with Alexa⁴⁸⁸ (1:400, Molecular Probes). The eye imaginal discs were

then mounted in Vectashield (Vector) and stored at -20°C until viewing. An image series of 40 focal planes (total thickness $16\text{ }\mu\text{M}$) was collected from the central region of each imaginal disc (384×384 pixel) with an Applied Precision Restoration Microscopy Optical Workstation (Applied Precision, Inc., Issaquah, WA). Optimal exposures were determined empirically to obtain images that did not contain any saturated pixels (that is, less than 4,095), and this exposure was used for all samples. Z-stacks (merged images) were then created by a constrained iterative algorithm using SoftWorx software (Applied Precision, Inc.). The total value of all pixels in the region was calculated by the software. To compare the values of different genotypes or temperature conditions, relative percentages shown in figures were calculated by dividing the higher value with the lower value in each data set.

Analysis of P/EP elements

Genomic DNA regions flanking the P and EP elements were recovered from modifier lines by standard plasmid rescue and/or inverse polymerase chain reaction (PCR) protocols (<http://www.fruitfly.org>). Recovered DNA was sequenced and analysed with BLAST and GENEFINDER (<http://dot.imgen.bcm.tmc.edu:9331/gene-finder>) in BDGP (<http://www.fruitfly.org>) and NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) to confirm identity of the P/EP elements and to identify the modifier gene.

To verify specific gene overexpression in EP elements, we obtained coding sequences downstream of the EP region by PCR and used them as probes for *in situ* hybridization in larvae carrying the *dppGal4* driver and the EP insertion of interest.

Received 16 June; accepted 4 October 2000.

- Cummings, C. J. & Zoghbi, H. Y. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **9**, 909–916 (2000).
- Brand, A. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Warrick, J. M. *et al.* Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70. *Nature Genet.* **23**, 425–428 (1999).
- Jackson, G. R. *et al.* Polyglutamine-expanded human huntingtin transgenes induce degeneration of *Drosophila* photoreceptor neurons. *Neuron* **21**, 633–642 (1998).
- Kazemi-Esfarjani, P. & Benzer, S. Genetic suppression of polyglutamine toxicity in *Drosophila*. *Science* **287**, 1837–1840 (2000).
- Marsh, J. L. *et al.* Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in *Drosophila*. *Hum. Mol. Genet.* **9**, 13–25 (2000).
- Lin, X., Cummings, C. J. & Zoghbi, H. Y. Expanding our understanding of polyglutamine diseases through mouse models. *Neuron* **24**, 499–502 (1999).
- Moses, K. & Rubin, G. M. Glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing *Drosophila* eye. *Genes Dev.* **5**, 583–593 (1991).
- Burridge, E. N. *et al.* SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* **82**, 937–948 (1995).
- Klement, I. A. *et al.* Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **95**, 41–53 (1998).
- Cummings, C. J. *et al.* Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. *Neuron* **24**, 879–892 (1999).
- Rorth, P. *et al.* Systematic gain-of-function genetics in *Drosophila*. *Development* **125**, 1049–1057 (1998).
- Chai, Y., Koppenhafer, S. L., Bonini, N. M. & Paulson, H. L. Analysis of the role of heat shock protein (Hsp) molecular chaperones in polyglutamine disease. *J. Neurosci.* **19**, 10338–10347 (1999).
- Salinas, A. E. & Wong, M. G. Glutathione S-transferases—a review. *Curr. Med. Chem.* **6**, 279–309 (1999).
- Bodoor, K. *et al.* Function and assembly of nuclear pore complex proteins. *Biochem. Cell Biol.* **77**, 321–329 (1999).
- Grams, R. & Korge, G. The *mub* gene encodes a protein containing three KH domains and is expressed in the mushroom bodies of *Drosophila melanogaster*. *Gene* **215**, 191–201 (1998).
- Lin, X., Antalfi, B., Kang, D., Orr, H. T. & Zoghbi, H. Y. Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1. *Nature Neurosci.* **3**, 157–163 (2000).
- Waragai, M. *et al.* PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum. Mol. Genet.* **8**, 977–987 (1999).
- Boutell, J. M. *et al.* Aberrant interactions of transcriptional repressor proteins with the Huntington's disease gene product, huntingtin. *Hum. Mol. Genet.* **8**, 1647–1655 (1999).
- Masliha, E. *et al.* Dopaminergic loss and inclusion body formation in alpha-synuclein mice: implications for neurodegenerative disorders. *Science* **287**, 1265–1269 (2000).
- Harel, A. *et al.* Persistence of major nuclear envelope antigens in an envelope-like structure during mitosis in *Drosophila melanogaster* embryos. *J. Cell Sci.* **94**, 463–470 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Mancini for help with the confocal microscope; K.-W. Choi and members of his laboratory for help with the eye sections; C. Cummings for advice and insightful discussions; V. Brandt for editorial help; M. Magarinos and P. Herrero for help with the initial fly screens; R. Davis and A. L. Beaudet for reading the manuscript; C. Cater, G. Rubin, D. Cribbs, R. Davis, T. Aigaki, H. Steller, G. Pennetta, S. Parkhurst and the Bloomington Stock Center for fly strains; and M. Levine and R. Wharton for antibodies. This work was supported by a grant of the NIH to J.B. J.B. is also grateful for initial support by the Banco Bilbao Vizcaya. M.L.N.-R. was supported by an HHMI postdoctoral fellowship for physicians. H.Y.Z. is a Howard Hughes Medical Institute Investigator.

Correspondence and requests for materials should be addressed to (e-mail: jbotas@bcm.tmc.edu).

Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation

Timothy A. McKinsey, Chun-Li Zhang, Jianrong Lu & Eric N. Olson

Department of Molecular Biology, The University of Texas Southwestern Medical Center at Dallas, 6000 Harry Hines Blvd, Dallas, Texas 75390-9148, USA

Members of the myocyte enhancer factor-2 (MEF2) family of transcription factors associate with myogenic basic helix–loop–helix transcription factors such as MyoD to activate skeletal myogenesis¹. MEF2 proteins also interact with the class II histone deacetylases HDAC4 and HDAC5, resulting in repression of MEF2-dependent genes^{2–4}. Execution of the muscle differentiation program requires release of MEF2 from repression by HDACs, which are expressed constitutively in myoblasts and myotubes⁵. Here we show that HDAC5 shuttles from the nucleus to the cytoplasm when myoblasts are triggered to differentiate. Calcium/calmodulin-dependent protein kinase (CaMK) signalling, which stimulates myogenesis⁵ and prevents formation of MEF2–HDAC complexes⁴, also induces nuclear export of HDAC4 and HDAC5 by phosphorylation of these transcriptional repressors. An HDAC5 mutant lacking two CaMK phosphorylation sites is resistant to CaMK-mediated nuclear export and acts as a dominant inhibitor of skeletal myogenesis, whereas a cytoplasmic HDAC5 mutant is unable to block efficiently the muscle differentiation program. Our results highlight a mechanism for transcriptional regulation through signal- and differentiation-dependent nuclear export of a chromatin-remodelling enzyme, and suggest that nucleo-cytoplasmic trafficking of HDACs is involved in the control of cellular differentiation.

Histone acetylation/deacetylation has emerged as a fundamental mechanism for the control of gene expression⁶. Histone acetyltransferases (HATs) stimulate transcription through acetylation of histones, resulting in relaxation of nucleosomes; and HDACs antagonize this activity and repress transcription⁶. HDAC4 and HDAC5 block myogenesis by associating with and inhibiting the activity of the MEF2 transcription factor, and CaMK signalling overcomes this inhibition by dissociating MEF2–HDAC complexes, with consequent stimulation of myogenesis^{4,5}.

To investigate further the mechanism for suppression of myogenesis by HDAC4 and HDAC5, and the blockade to this suppression by CaMK signalling, we raised antibodies against HDAC5 and examined the subcellular distribution of the endogenous protein during differentiation of the C2 skeletal muscle cell line. In proliferating undifferentiated myoblasts, HDAC5 was predominantly nuclear (Fig. 1a), whereas after initiation of differentiation by removal of serum from the medium, HDAC5 became progressively localized in the cytoplasm, although residual nuclear staining was evident (Fig. 1b–d). In contrast, MEF2 was exclusively nuclear in myotubes (Fig. 1e), consistent with the essential role for this transcription factor in skeletal muscle gene expression. In residual unfused myoblasts that failed to differentiate, HDAC5 remained in the nucleus, suggesting that cytoplasmic accumulation of HDAC5 is coupled to activation of the muscle differentiation program.

As activated CaMK can overcome HDAC-mediated repression of myogenesis⁵, we investigated whether CaMK signalling also alters the subcellular distribution of HDAC5. In the absence of activated CaMK, epitope-tagged derivatives of MEF2C and HDAC5 were coexpressed exclusively in the nuclei of Cos cells (Fig. 2a). In contrast, HDAC5 was cytoplasmic in cells expressing a constitutively active form of CaMKI. MEF2C remained nuclear in CaMKI-expressing cells, consistent with the finding that CaMK signalling

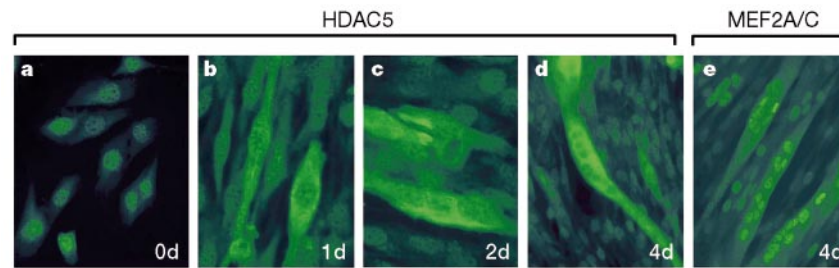


Figure 1 Shuttling of HDAC5 from the nucleus to the cytoplasm during myogenic differentiation. Undifferentiated C2 myoblasts maintained in growth medium (**a**), and cells exposed to differentiation medium for one (**b**), two (**c**) or four (**d**) days were stained with antibodies directed against the N terminus of HDAC5 and a fluorescein-conjugated secondary antibody. We note that HDAC5 staining remains mostly nuclear in myoblasts

that fail to differentiate into myotubes, whereas it is mostly cytoplasmic in adjacent myotubes. Myotubes were also stained with antiserum that recognizes both MEF2A and MEF2C (**e**). Images in **b** and **c** were magnified about 1.5 times relative to those in **a**, **d** and **e** to enable better visualization of elongated cells in the early stages of differentiation.

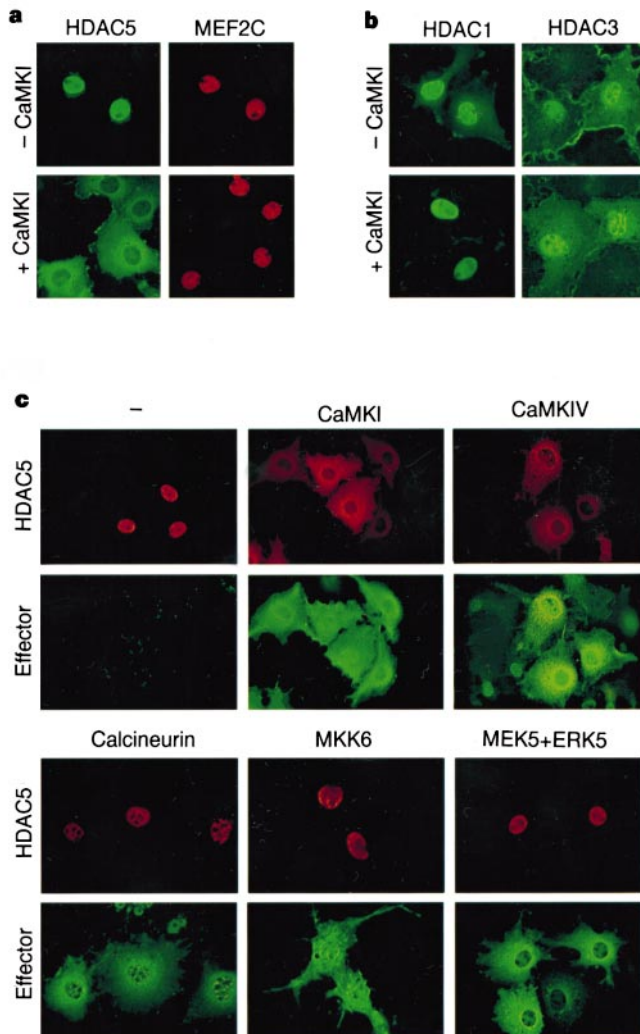


Figure 2 HDAC5 is excluded from the nucleus in cells expressing activated forms of CaMK. **a**, **b**, Cos cells were co-transfected with expression vectors encoding Flag-tagged HDAC5 and Myc-tagged MEF2C (**a**) or Flag-tagged HDAC1 or HDAC3 (**b**) in the absence or presence of a plasmid for activated CaMKI. HDACs and MEF2C were detected by indirect immunofluorescence using antibodies against the epitope tags and fluorescein-conjugated (HDACs) and rhodamine-conjugated (MEF2C) secondary antibodies. **c**, Cos cells were co-transfected with expression vectors for Myc-tagged HDAC5 and epitope-tagged derivatives of the indicated MEF2 activators. Cells were analysed by indirect immunofluorescence using antibodies against the epitope tags and rhodamine-conjugated (HDAC5) and fluorescein-conjugated (Effector) secondary antibodies. ERK5 staining is shown for cells co-expressing MEK5 and ERK5.

disrupts MEF2–HDAC complexes and stimulates MEF2-dependent transcription⁴.

Association of HDAC5 with MEF2 is mediated by an 18-amino-acid motif near the amino terminus^{2–4} (Fig. 3e). A similar region is contained in HDAC4, but not in the class I HDAC1 or HDAC3. Nuclear exclusion in response to CaMK signalling was also observed with HDAC4 (data not shown), but not with HDAC1 and HDAC3 (Fig. 2b). Identical results were obtained in Cos, HeLa, 293 and 10T1/2 fibroblasts, as well as in C2 and L6 myoblasts (data not shown); results with Cos cells are shown because they express the exogenous proteins at highest levels.

In addition to CaMK, mitogen-activated protein kinases such as MKK6 and MEK5, which stimulate phosphorylation of the MEF2 transactivation domain through p38 (ref. 7) and ERK5 (ref. 8), respectively, activate MEF2. The calcium/calmodulin-dependent phosphatase calcineurin also stimulates MEF2 activity^{9,10}. We therefore investigated whether these MEF2 activators also promote cytoplasmic accumulation of HDAC5. CaMKIV, like CaMKI, caused nuclear exclusion of HDAC5, whereas activated forms of MKK6, ERK5 and calcineurin had no effect on HDAC5 nuclear localization (Fig. 2c), further establishing the specificity of this signalling process.

Nuclear exclusion of HDAC5 could result from inhibition of nuclear import or stimulation of nuclear export. To distinguish between these possibilities, the subcellular distribution of an HDAC5–green fluorescent protein (GFP) chimaeric molecule was monitored in cells exposed to leptomycin B, a fungal toxin that blocks nuclear export by the exportin protein Crm1 (refs 11, 12). HDAC5–GFP normally resides in the nucleus, but is mostly cytoplasmic in cells expressing activated CaMKI (Fig. 3a). Treatment of CaMKI-expressing cells with leptomycin B resulted in translocation of HDAC5 from the cytoplasm to the nucleus over a time course of 4 h (Fig. 3a). Similar results were obtained in cells translationally arrested with cycloheximide (data not shown), indicating that newly synthesized, as well as pre-existing, cytoplasmic HDAC5 enters the nucleus after leptomycin B treatment. These results indicate that CaMK signalling stimulates nuclear export of HDAC5 through a Crm1-dependent mechanism.

Protein kinase A (PKA) and glycogen synthase kinase-3 (GSK-3) promote nuclear export of the Msn2 (ref. 13) transcription factor and two members of the NFAT family¹⁴ of transcription factors, respectively, and the AKT kinase stimulates nuclear export of the forkhead transcription factor¹⁵. Of note, the consensus phosphorylation sites for PKA and AKT (R/K-x-x-S/T) overlap precisely with that of CaMK¹⁶. Despite their overlapping substrate specificity, however, neither PKA, AKT nor GSK-3 was capable of driving HDAC5 from the nucleus to the cytoplasm (Fig. 3b).

To identify sequences in HDAC5 that control nuclear localization and responsiveness to CaMK signalling, we generated serial car-

boxy-terminal truncation mutants of HDAC5 and examined them by indirect immunofluorescence. Full-length HDAC5 (amino acids 1–1122) was localized exclusively in the nuclei of transfected Cos cells (Fig. 3c). However, removal of about 100 C-terminal amino acids from HDAC5 (mutant 1–1021) led to a diffuse, whole-cell pattern of HDAC5 localization, and a mutant lacking an additional 100 amino acids (mutant 1–921) was completely excluded from the

nucleus. Further deletion of HDAC5 C-terminal sequences up to residue 360 resulted in a nuclear pattern of HDAC5 localization, whereas a mutant comprising the first 260 amino acids of HDAC5 (mutant 1–260) was primarily cytoplasmic. These results suggest that sequences in the N and C termini of HDAC5 regulate nuclear import and export, respectively (see Fig. 3f). Consistent with these predictions, an HDAC5 mutant lacking amino acids 260–304 was

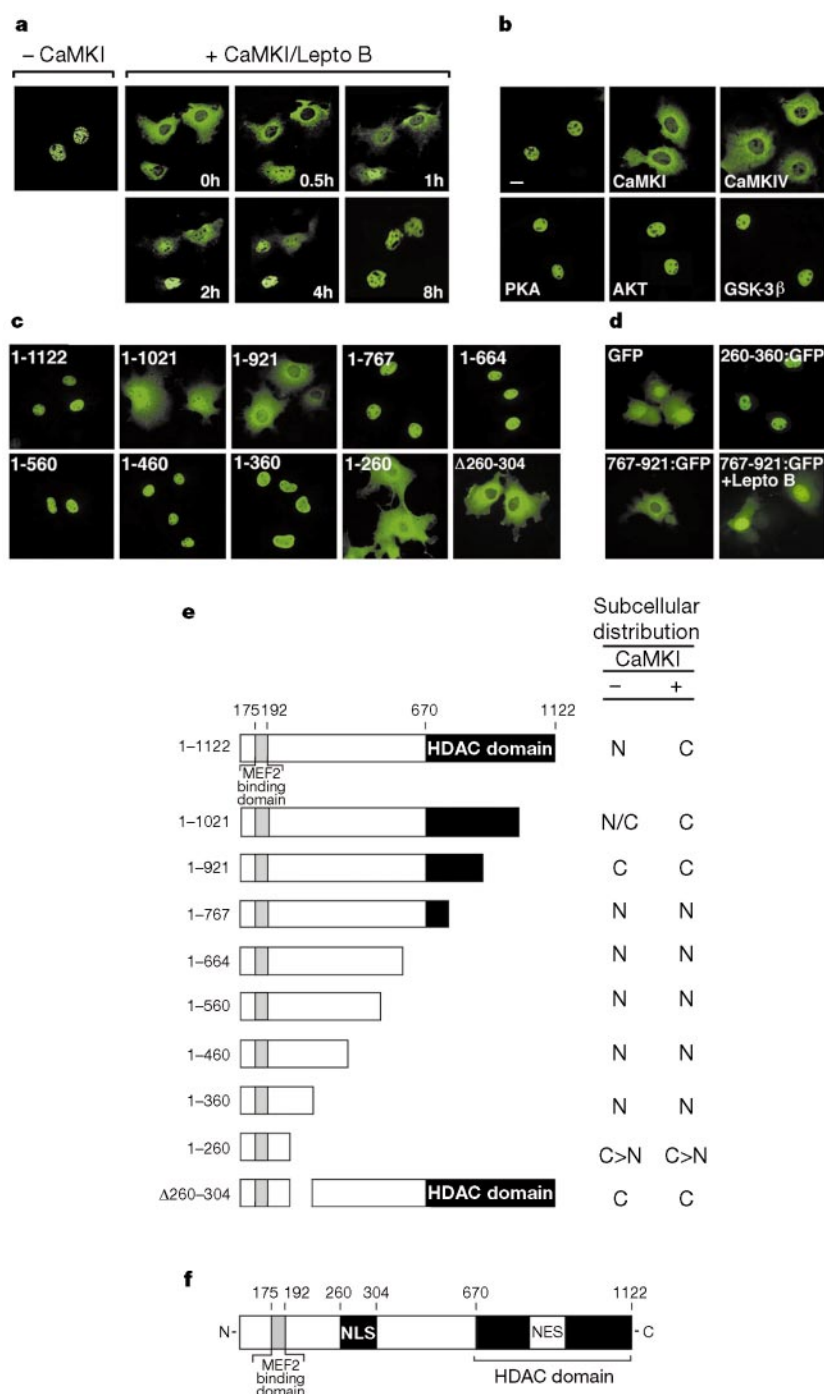


Figure 3 CaMK signalling stimulates nuclear export of HDAC5. **a**, Cos cells were transfected with an expression vector encoding HDAC5 fused to GFP in the absence or presence of a vector for activated CaMKI. GFP-positive cells were photographed before and after addition of leptomycin B (20 ng ml⁻¹) to the medium for the indicated times. **b**, **c**, Cos cells were transfected with expression vectors for Flag-tagged HDAC5 and activated AKT, wild-type PKA, or wild-type GSK-3β (**b**) or Flag-tagged HDAC5 deletion mutants (**c**). HDAC5 was detected by indirect immunofluorescence using anti-Flag

antibodies and a fluorescein-conjugated secondary antibody. **d**, Cos cells were transfected with expression vectors for the indicated HDAC5–GFP fusion proteins. GFP-positive cells were photographed 24 h after transfection. Leptomycin B (20 ng ml⁻¹) was added to HDAC5(767–921)–GFP expressing cells 8 h before analysis. **e**, Subcellular distribution of HDAC5 mutants in the absence and presence of activated CaMK. C, cytoplasmic; N, nuclear; N/C, whole-cell; C>N, cytoplasmic greater than nuclear. **f**, Diagram of HDAC5.

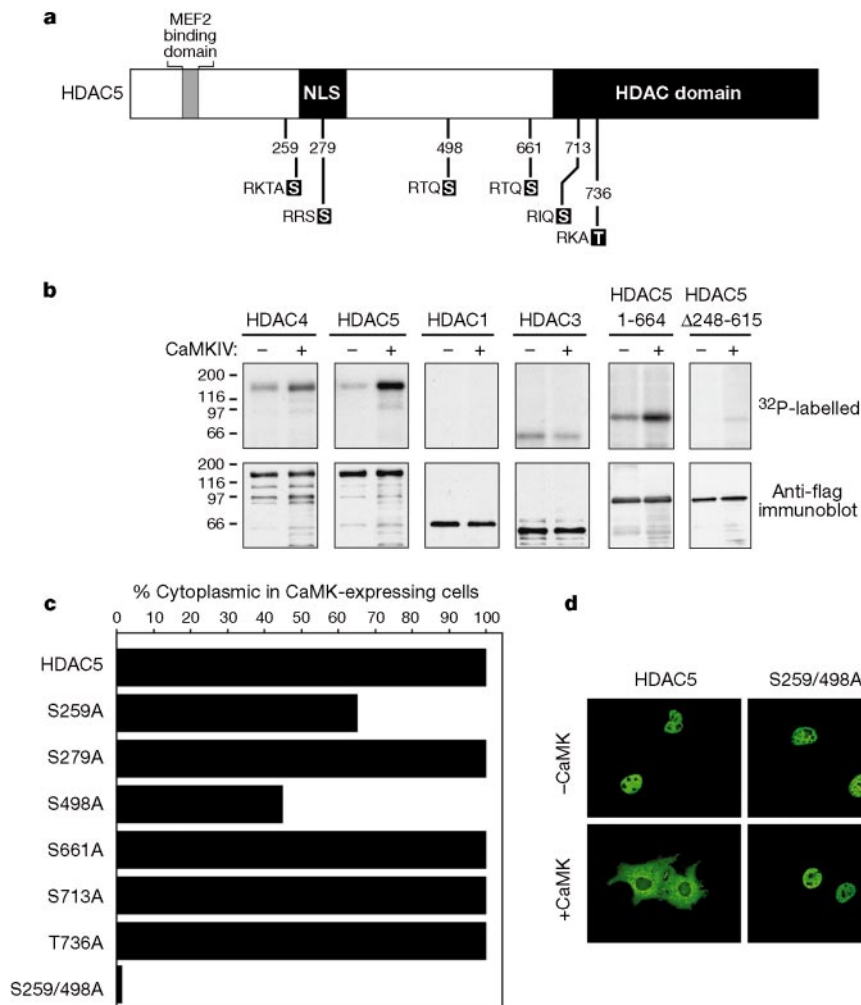


Figure 4 Identification of CaMK target sites in HDAC5. **a**, Six consensus CaMK sites in HDAC5 that are conserved in HDAC4 are shown. **b**, Cos cells were transfected with expression vectors encoding Flag-tagged forms of the indicated HDAC or a Flag-tagged derivative of activated CaMKIV and Flag-tagged proteins were immunoprecipitated and incorporated into *in vitro* phosphorylation reactions (see Methods). Proteins were resolved by SDS-PAGE, transferred to PVDF membranes, and visualized by autoradiography (top) followed by immunoblotting with anti-Flag antibodies (bottom). **c**, Cos cells were co-

transfected with expression vectors for Flag-tagged forms of the indicated HDAC5 point mutants and an expression plasmid for an HA-tagged derivative of activated CaMKI. Cells were analysed by double immunofluorescence using a monoclonal anti-Flag antibody and polyclonal antibodies against the HA tag. Values represent the percentage of CaMKI-expressing cells in which HDAC5 exhibited cytoplasmic staining. In **d**, the immunofluorescence images reveal the loss of CaMK responsiveness of the HDAC5(S259/498A) double mutant.

cytoplasmic (Fig. 3c; Δ260-304), and all of the HDAC5 mutants lacking sequences C-terminal to amino acid 767 remained predominantly nuclear in the presence of activated CaMKI (Fig. 3e; data not shown).

To assess whether the above domains in HDAC5 act as true nuclear import and export signals, we fused amino acids 260–360 and 761–921 of HDAC5 to the N terminus of GFP and determined the subcellular distribution of the resultant fusion proteins. Unmodified GFP was localized in both the nucleus and cytoplasm of transfected cells (Fig. 3d). When fused to amino acids 260–360 of HDAC5, GFP was driven into the nucleus. In contrast, a GFP fusion protein containing amino acids 761–921 of HDAC5 was exclusively cytoplasmic, and partially relocated to the nucleus in the presence of leptomycin B. These results show that specific portions of HDAC5 can act as nuclear import or export signals when fused to a heterologous protein.

HDAC5 contains six consensus CaMK phosphorylation sites that are conserved in HDAC4 (Fig. 4a). To determine whether HDACs can act as direct substrates for CaMK, we immunoprecipitated ectopic HDACs from transfected Cos cells and incorporated them

into reaction mixtures containing [γ - 32 P]ATP in the absence or presence of activated CaMKIV. HDAC4, -5 and -3 were phosphorylated in the absence of exogenous CaMK (Fig. 4b), indicating that these HDACs may associate stably with one or more endogenous kinases present in Cos cells. In the presence of activated CaMKIV, the level of phosphorylation of HDAC4 and -5 was significantly increased, whereas the phosphorylation status of HDAC1 and -3 was unchanged. Analysis of deletion mutants localized the CaMK phosphorylation sites to a segment of the N-terminal extension of HDAC5 (amino acids 248–615), which contains three consensus CaMK phosphorylation sites: Ser 259, Ser 279 and Ser 498 (see Fig. 4a). Similar sites influence constitutive cytoplasmic localization of HDAC4 in U2OS cells¹⁷.

If the subcellular distribution of HDAC5 is governed by its phosphorylation status, then disruption of the relevant phosphoacceptor sites would be predicted to block CaMK-mediated nuclear export. To test this hypothesis, each potential CaMK phosphorylation site in HDAC5 was substituted with alanine and the resultant mutants were examined for their ability to undergo nuclear export in response to CaMK signalling. In agreement with the *in vitro*

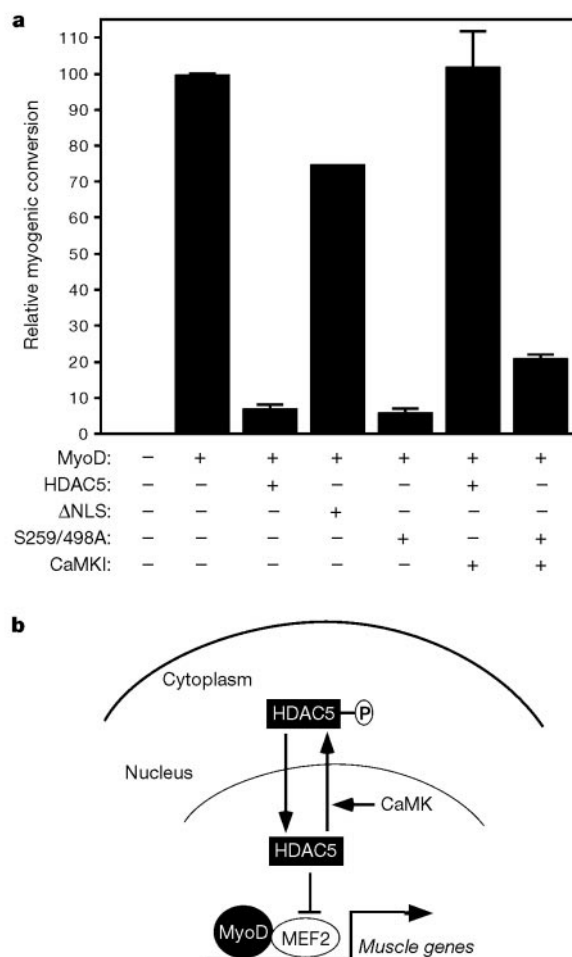


Figure 5 Regulation of myogenesis by HDAC5 nuclear export. **a**, 10T1/2 fibroblasts were co-transfected with expression vectors for MyoD and the indicated HDAC5 protein in the absence or presence of a plasmid for activated CaMKI. Cells were transferred to differentiation medium 2 d post-transfection and stained with anti-myosin antibodies after 4 d more in culture. Relative myogenic conversion represents the percentage of myosin-positive cells in HDAC transfectants relative to cells expressing MyoD alone (100 represents about 500 myosin-positive cells per 35-mm dish). Values represent the mean $\pm$ s.d. from at least two experiments. **b**, Model for signal-dependent regulation of myogenesis. HDAC5 blocks muscle differentiation by repressing the transcriptional activity of MEF2. CaMK phosphorylates HDAC5 and stimulates its nuclear export, freeing MEF2 to cooperate with MyoD to activate genes required for skeletal myogenesis.

phosphorylation data (Fig. 4b), mutation of Ser 661, Ser 713 or Thr 736 in HDAC5 had no effect on CaMK-dependent nuclear export (Fig. 4c). Likewise, disruption of Ser 279, which lies within the region of HDAC5 phosphorylated by CaMK (see Fig. 4b), did not prevent nuclear export in the presence of activated CaMK. In contrast, replacement of either Ser 259 or Ser 498 with alanine markedly reduced CaMK-mediated trafficking of HDAC5 from the nucleus to the cytoplasm, indicating that these sites may be involved in nuclear export. Indeed, simultaneous disruption of Ser 259 and Ser 498 completely blocked CaMK-mediated nuclear export of HDAC5 (Fig. 4c, d) and reduced the degree of phosphorylation by CaMK (data not shown). This mutant was still showed low-level phosphorylation, however, suggesting that phosphorylation at other sites in HDAC5 is not sufficient to induce nuclear export. These results show that phosphorylation of Ser 259 and Ser 498 is required for nuclear export of HDAC5.

To address the role of HDAC5 nuclear export in muscle differentiation, we tested the HDAC5(S259/498A) mutant for its effects

on MyoD-dependent muscle gene activation in 10T1/2 fibroblasts, which requires MEF2 (ref. 1) and HAT activity¹⁸. Transfection of 10T1/2 cells with MyoD alone resulted in efficient conversion to the muscle cell lineage, assayed by myosin staining (Fig. 5a). Overexpression of wild-type HDAC5 interfered with this process. The HDAC5(S259/498A) mutant consistently blocked MyoD-dependent myogenic conversion of 10T1/2 cells even more potently than wild-type HDAC5. In contrast, a constitutively cytoplasmic mutant of HDAC5 lacking its nuclear localization signal (Δ NLS) was severely impaired in its ability to inhibit myogenesis. As described previously⁵, CaMK signalling completely rescued cells from the block to myogenesis imposed by HDAC5 (Fig. 5a). In cells expressing the HDAC5(S259/498A) mutant, however, the myogenic program remained largely blocked, despite the presence of activated CaMK. Together, these results suggest that nuclear HDAC5 represses myogenesis and that export of HDAC5 from the nucleus is required to execute the myogenic program.

A model for signal-dependent regulation of myogenesis through nuclear export of HDAC5 is depicted in Fig. 5b. Although our results show the potency and specificity of CaMK for promoting nuclear export of HDAC5, it remains possible that other signalling molecules with substrate specificities similar to that of CaMK can regulate HDAC5 nuclear export in certain cellular contexts. In addition to its apparent role in skeletal myogenesis, CaMK signalling regulates learning and memory¹⁹ as well as cardiac hypertrophy²⁰. As class II HDACs and MEF2 are expressed primarily in skeletal muscle, heart and brain^{1,21}, it is intriguing to speculate that CaMK controls these diverse biological processes by stimulating nuclear export of HDACs, resulting in activation of MEF2-dependent genes. □

Methods

Cell culture

Cos, C2 and 10T1/2 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with fetal bovine serum (10%), L-glutamine and penicillin/streptomycin. Lipid-mediated DNA transfections were performed using Fugene 6 reagent (Roche Molecular Biochemicals) and 0.5 μ g of the indicated plasmids.

Myogenic conversion assays

For myogenic differentiation of confluent C2 and MyoD-transfected 10T1/2 cultures, growth medium was replaced with DMEM containing horse serum (2%), L-glutamine and penicillin/streptomycin. Transiently transfected 10T1/2 cells were exposed to differentiation medium 2 d after transfection and assessed by immunoperoxidase staining with the Vectastain Elite ABC kit (Vector Laboratories), and an anti-skeletal myosin antibody (Sigma) after a further 4 d.

Plasmids

Mammalian expression vectors for HDAC1, HDAC3 and HDAC5 (all with C-terminal Flag tags)²², constitutively active mutants of CaMKI²³, CaMKIV (C-terminal Flag tag)²⁴, calcineurin²⁵, MKK6²⁶, AKT²⁷, MEK5²⁸, ERK5²⁸ and wild-type forms of GSK-3 β (ref. 29) and PKA (ref. 30) have been described. The cDNAs encoding activated CaMKI²³ and CaMKIV²⁴ contain stop codons in place of Ile 294 and Gln 318, respectively, thereby removing the C-terminal autoinhibitory domains. Both of these CaMK mutants function constitutively without the requirement for calcium and calmodulin for activation. We fused sequences encoding the indicated epitope tags to MEF2C (C-terminal Myc tag), HDAC5 (N-terminal Myc tag), calcineurin (N-terminal Flag tag), MKK6 (C-terminal Flag tag) and CaMKI (N-terminal 3 $\times$ HA (haemagglutinin) tag), in the pCDNA3.1 expression vector (Invitrogen). HDAC5-GFP fusion proteins were constructed in the pEGFP-N1 expression plasmid (Clontech). Site-directed mutagenesis was performed using the QuikChange kit (Stratagene). For mutants S279A and S661A, the adjacent serines at positions 278 and 662, respectively, were also converted to alanine.

Indirect immunofluorescence

Cos cells were grown on glass coverslips, fixed in 10% paraformaldehyde, and stained in PBS containing 3% bovine serum albumin and 0.1% nonidet-P40. C2 cells were grown on gelatin-coated glass coverslips, fixed in methanol and stained in PBS containing 3% goat serum and 0.1% Nonidet-P40. Primary antibodies against Flag (mouse monoclonal; Sigma), Myc (rabbit polyclonal; Santa Cruz) and HA (rabbit polyclonal; Santa Cruz) were used at a dilution of 1:200. Secondary fluorochrome-conjugated antibodies (Vector Labs) were also used at a dilution of 1:200. Rabbit polyclonal antisera for HDAC5 were raised against a glutathione-S-transferase fusion protein containing amino acids 11–70 of rat HDAC5. For immunostaining with anti-HDAC5 antibodies, we purified IgG fractions from crude serum using protein A-agarose beads (Pierce).

In vitro phosphorylation assays

For *in vitro* phosphorylation assays, Cos cells were transiently transfected with expression vectors encoding Flag-tagged HDACs or a Flag-tagged derivative of activated CaMKIV. Cells were collected in PBS containing 0.5% Triton X-100, 1 mM EDTA, 1 mM sodium pyrophosphate, 2 mM sodium fluoride, 10 mM β -glycerol phosphate, 1 mM sodium molybdate, 1 mM sodium orthovanadate, 1 mM PMSF and a protease inhibitor cocktail (Complete; Roche Molecular Biochemicals). Cells were sonicated briefly and cellular debris was removed by centrifugation. Flag-tagged proteins were immunoprecipitated using anti-Flag affinity resin (Sigma). Immunoprecipitates were washed two times in lysis buffer followed by three additional washes in a solution of 25 mM HEPES pH 7.6 and 10 mM MgCl_2 . Where indicated, CaMKIV-bound beads were mixed with those bound to HDAC immediately before the addition of reaction buffer (40 μl) consisting of 25 mM HEPES pH 7.6, 10 mM MgCl_2 , 2 mM CaCl_2 , 20 $\mu\text{g ml}^{-1}$ calmodulin, 12.5 μM ATP and [γ - ^{32}P]ATP (10 μCi ; 4,500 Ci mmol^{-1}). Reaction mixtures were incubated at 25 °C for 20 min. Reactions were terminated by addition of SDS–PAGE loading buffer followed by boiling. Immunoprecipitated proteins were resolved by SDS–PAGE, transferred to PVDF membranes and visualized by autoradiography followed by immunoblotting with anti-Flag antibodies.

Received 31 July; accepted 29 August 2000.

1. Molkenin, J. D., Black, B. L., Martin, J. F. & Olson, E. N. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83**, 1125–1136 (1995).
2. Sparrow, D. B. *et al.* MEF-2 function is modified by a novel co-repressor, MITR. *EMBO J.* **18**, 5085–5098 (1999).
3. Miska, E. A. *et al.* HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J.* **18**, 5099–5107 (1999).
4. Lu, J., McKinsey, T. A., Nicol, R. L. & Olson, E. N. Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases. *Proc Natl Acad Sci. USA* **97**, 4070–4075 (2000).
5. Lu, J., McKinsey, T. A., Zhang, C. L. & Olson, E. N. Regulation of skeletal myogenesis by association of MEF2 with class II histone deacetylases. *Mol. Cell* **6**, 233–244 (2000).
6. Kuo, M. H. & Allis, C. D. Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20**, 615–626 (1998).
7. Han, J., Jiang, Y., Li, Z., Kravchenko, V. V. & Ulevitch, R. J. Activation of the transcription factor MEF2C by the MAP kinase p38 in inflammation. *Nature* **386**, 296–299 (1997).
8. Kato, Y. *et al.* BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. *EMBO J.* **16**, 7054–7066 (1997).
9. Mao, Z. & Wiedmann, M. Calcineurin enhances MEF2 DNA binding activity in calcium-dependent survival of cerebellar granule neurons. *J. Biol. Chem.* **274**, 31102–31107 (1999).
10. Wu, H. *et al.* MEF2 responds to multiple calcium-regulated signals in the control of skeletal muscle fiber type. *EMBO J.* **19**, 1–11 (2000).
11. Nishi, K. *et al.* Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. *J. Biol. Chem.* **269**, 6320–6324 (1994).
12. Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–311 (1997).
13. Gerner, W. *et al.* Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes Dev.* **12**, 586–597 (1998).
14. Beals, C. R., Sheridan, C. M., Turck, C. W., Gardner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
15. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
16. Pinna, L. A. & Ruzzene, M. How do protein kinases recognize their substrates? *Biochem. Biophys. Acta* **1314**, 191–225 (1996).
17. Grozinger, C. M. & Schreiber, S. L. Regulation of histone deacetylase 4 and 5 and transcriptional activity by 14-3-3-dependent cellular localization. *Proc. Natl Acad. Sci. USA* **97**, 7835–7840 (2000).
18. Sartorelli, V., Huang, J., Hamamori, Y. & Kedes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* **17**, 1010–1026 (1997).
19. Mayford, M. *et al.* Control of memory formation through regulated expression of a CaMKII transgene. *Science* **274**, 1678–1683 (1996).
20. Passier, R. *et al.* CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. *J. Clin. Invest.* **105**, 1395–1406 (2000).
21. Black, B. L. & Olson, E. N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Biol.* **14**, 167–196 (1998).
22. Grozinger, C. M., Hassig, C. A. & Schreiber, S. L. Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proc. Natl Acad. Sci. USA* **96**, 4868–4873 (1999).
23. Haribabu, B. *et al.* Human calcium-calmodulin dependent protein kinase I: cDNA cloning, domain structure and activation by phosphorylation at threonine-177 by calcium-calmodulin dependent protein kinase I kinase. *EMBO J.* **14**, 3679–3686 (1995).
24. Chatila, T., Anderson, K. A., Ho, N. & Means, A. R. A unique phosphorylation-dependent mechanism for the activation of Ca^{2+} /calmodulin-dependent protein kinase type IV/GR. *J. Biol. Chem.* **271**, 21542–21548 (1996).
25. O'Keefe, S. J., Tamura, J., Kincaid, R. L., Tocci, M. J. & O'Neill, E. A. FK-506- and CsA-sensitive activation of the interleukin-2 promoter by calcineurin. *Nature* **357**, 692–694 (1992).
26. Jiang, Y. *et al.* Characterization of the structure and function of a new mitogen-activated protein kinase (p38beta). *J. Biol. Chem.* **271**, 17920–17926 (1996).
27. Rybkin, I. L., Cross, M. E., McReynolds, E. M., Lin, R. Z. & Ballou, L. M. α 1A adrenergic receptor induces eukaryotic initiation factor 4E-binding protein 1 phosphorylation via a Ca^{2+} -dependent pathway independent of phosphatidylinositol 3-kinase/Akt. *J. Biol. Chem.* **275**, 5460–5465 (2000).
28. English, J. M. *et al.* Contribution of the ERK5/MEK5 pathway to Ras/Raf signaling and growth control. *J. Biol. Chem.* **274**, 31588–31592 (1999).

29. Stambolic, V. & Woodgett, J. R. Mitogen inactivation of glycogen synthase kinase-3 beta in intact cells via serine 9 phosphorylation. *Biochem. J.* **303**, 701–704 (1994).
30. Mellon, P. L., Clegg, C. H., Correll, L. A. & McKnight, G. S. Regulation of transcription by cyclic AMP-dependent protein kinase. *Proc. Natl Acad. Sci. USA* **86**, 4887–4891 (1989).

Acknowledgements

We thank M. Cobb, J. Han, R. Lin, G.S. McKnight, A. Means, S. O'Keefe, S. Schreiber, T. Chatila and J. Woodgett for expression plasmids; R. Prives for anti-MEF2 antisera; and M. Yoshida for leptomycin B. We are grateful to A. Tizenor for graphics, and J. Page and W. Simpson for editorial assistance. E.N.O. was supported by grants from NIH, the Robert A. Welch Foundation, the D. W. Reynolds Foundation and Myogen, Inc. T.A.M. is a Pfizer fellow of The Life Sciences Research Foundation.

Correspondence and requests for materials should be addressed to E.N.O. (e-mail: eolson@hamon.swmed.edu).

Structural basis for signal transduction by the Toll/interleukin-1 receptor domains

Yingwu Xu*, Xiao Tao*, Baohe Shen*, Tiffany Horng†, Ruslan Medzhitov†, James L. Manley* & Liang Tong*

* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

† Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA

Toll-like receptors (TLRs) and the interleukin-1 receptor superfamily (IL-1Rs) are integral to both innate and adaptive immunity for host defence^{1–3}. These receptors share a conserved cytoplasmic domain^{4,5}, known as the TIR domain. A single-point mutation in the TIR domain of murine TLR4 (Pro712His, the Lps^d mutation) abolishes the host immune response to lipopolysaccharide (LPS)⁶, and mutation of the equivalent residue in TLR2, Pro681His, disrupts signal transduction in response to stimulation by yeast and Gram-positive bacteria⁷. Here we report the crystal structures of the TIR domains of human TLR1 and TLR2 and of the Pro681His mutant of TLR2. The structures have a large conserved surface patch that also contains the site of the Lps^d mutation. Mutagenesis and functional studies confirm that residues in this surface patch are crucial for receptor signalling. The Lps^d mutation does not disturb the structure of the TIR domain itself. Instead, structural and functional studies indicate that the conserved surface patch may mediate interactions with the downstream MyD88 adapter molecule^{7–11}, and that the Lps^d mutation may abolish receptor signalling by disrupting this recruitment.

A principal function of TIR domains is thought to be to mediate homotypic protein–protein interactions in the signal transduction process². To elucidate the molecular basis of TIR domain signalling, we have determined the crystal structures of the TIR domains of human TLR1 and TLR2 (Table 1). The structures contain a central five-stranded parallel β -sheet (β A– β E) that is surrounded by a total of five helices (α A– α E) on both sides (Fig. 1a). To facilitate sequence and structure comparisons of this large family of protein domains, we use a residue numbering system based on the secondary structure elements observed in the current structures, similar to that adopted for Src homology (SH)-2 domains¹². In the numbering system, a residue in a β -strand or α -helix is numbered according to its position in that strand or helix. The loops are named by the letters of the secondary structure elements that they connect. For example, the BB loop connects strand β B and helix α B. A residue in a loop is numbered according to its position in that loop, or

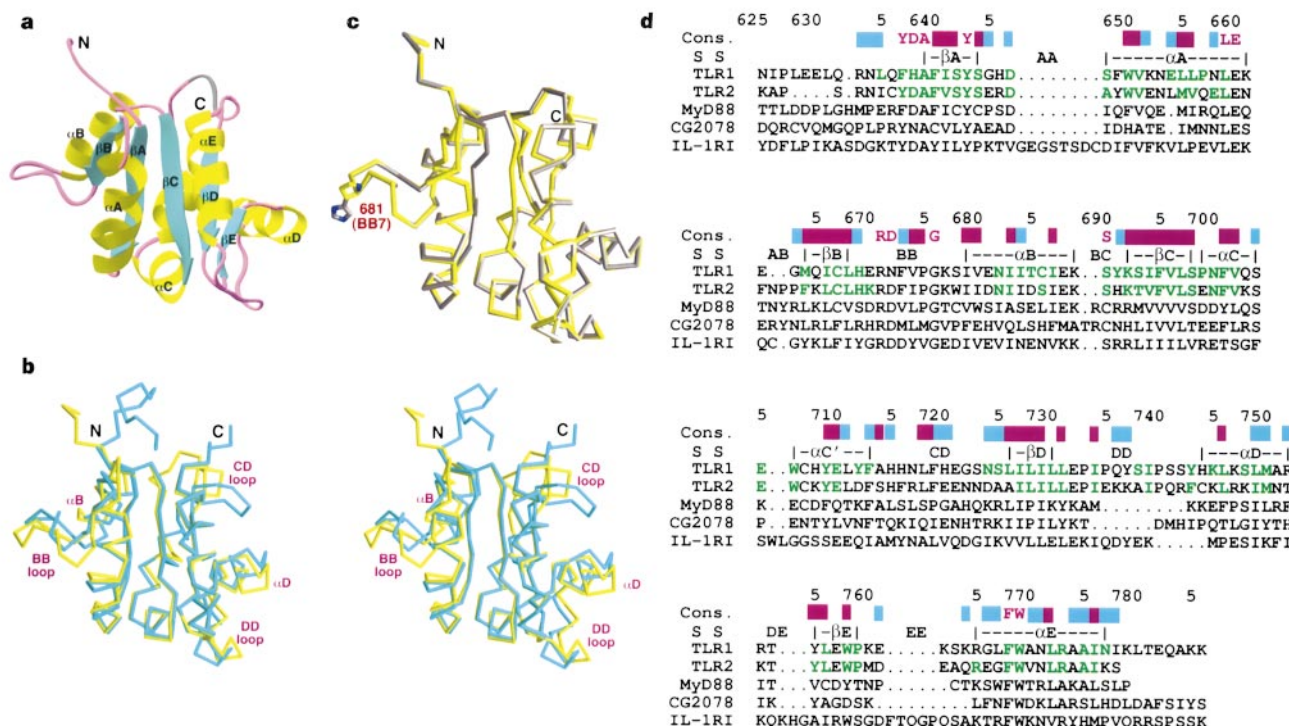


Figure 1 Structures of the TIR domain. **a**, Drawing of the structure of the TIR domain of human TLR2. Three missing residues, 723–725, are shown in grey. **b**, Superposition of the C α traces of the structures of the TIR domain of human TLR1 (cyan) and TLR2 (yellow). Regions of large differences between the two structures are labelled. **c**, Superposition of the C α traces of the structures of the TIR domain of human TLR2 wild type (yellow) and the P681H mutant (grey). The side chains of residue 681 are shown. **d**, Alignment of

according to its position relative to the nearest strand or helix. For example, the residue immediately preceding strand β A can be numbered as β A-1.

The structures reveal that the core TIR domain starts from the conserved (F/Y)DA amino-acid motif and ends roughly eight residues carboxy-terminal to the conserved FW motif (Fig. 1d). The alanine residue of the (F/Y)DA motif is the first residue in strand β A (the β A1 residue). The amino and carboxy termini of this core TIR domain are located within 14 Å of each other, suggesting that the TIR domain may be considered as a cassette. Most of the conserved residues are located in the hydrophobic core of the structure. There are large insertions or deletions in several loop regions in different TIR domains. Consequently, the sizes of the core TIR domains vary considerably—between 135 and 160 residues among the sequences reported so far.

Although the TIR domains of TLR1 and TLR2 share 50% amino-acid sequence identity, there are large conformational differences between the two structures (for example, helices α B, α C', and α D) (Fig. 1b). Noting that the sequence conservation among TIR domains is generally in the 20–30% range, as well as the variation in the sizes of the domains, our studies suggest that TIR domains may have a significant amount of sequence and structural diversity. This diversity may be crucial for the specificity in the signal transduction process, by ensuring the formation of the proper signalling complex among the many TLRs and IL-1Rs.

Three types of TIR domain interactions are possible for receptor signalling in animals. The first interface (which we call the R face) would mediate the oligomerization of receptor TIR domains, facilitated by the ligand-induced association of the extracellular domains of the receptors. The second interface (the A face, possibly equivalent to the R face in the receptor) would mediate the

oligomerization of the TIR domains of the downstream adapter molecule (MyD88), which may be facilitated by death domain interactions in this molecule. It has been shown that MyD88 exists as dimers in solution¹¹. The third interface (the S face) would mediate the association between the receptor and adapter TIR domains, and the formation of this TIR domain complex is critical for receptor signalling^{7–11}. In *Drosophila*, the gene CG2078 has the same domain organization as MyD88 and may therefore be a functional homologue of MyD88, suggesting that the TIR domain signalling complex may be conserved between insects and mammals.

The interactions that take place at the R face are probably a main determinant of specificity in the receptor signalling process, in addition to those between the ligand and the receptor. This is supported by the observation that chimaeric receptor containing the extracellular domain of IL-1R type I (IL-1RI) and the intracellular domain of TLR1 cannot function in response to IL-1 (ref. 13), a possible explanation being that the TIR domains of the IL-1R accessory protein and TLR1 cannot form a proper complex. Therefore, it is likely that interactions at the R face may show a significant degree of diversity to allow for the specificity, consistent with the sequence and structural diversity that we have observed. Residues at this interface are probably not strictly conserved among the various receptor TIR domains.

In contrast, interactions at the S face should be primarily mediated by a highly conserved region among the TIR domains. Despite the identification of more than 20 TLRs and IL-1Rs in mammals, only one putative adapter molecule (MyD88) is currently known to contain a TIR domain^{7–11}. Tollip has recently been identified as a receptor-proximal component of the IL-1RI pathway, but it does not contain a TIR domain¹⁴. In *Drosophila*, there are a

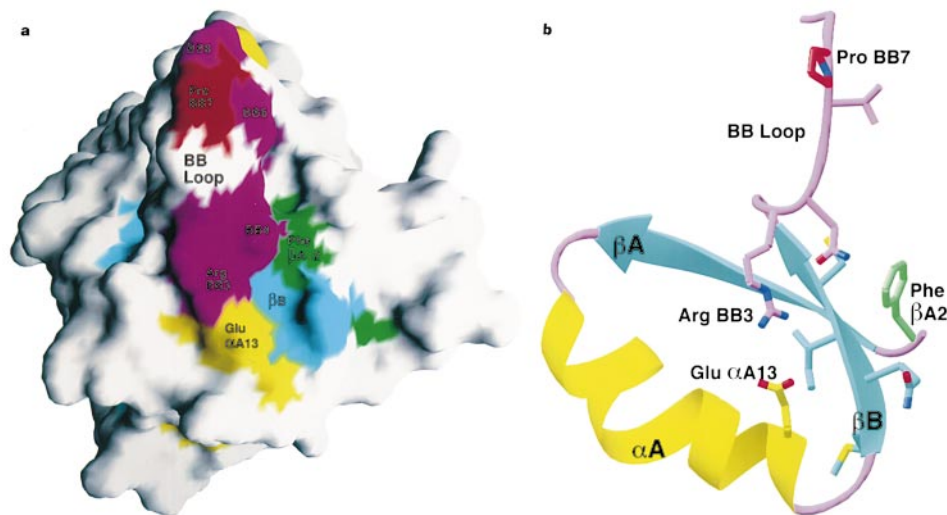


Figure 2 Structural analysis of the TIR domains. **a**, A large patch of conserved surface residues in the BB loop. Molecular surface of the TIR domain of TLR2 is shown. Conserved residues are highlighted in β -strands (cyan) in helices (yellow), in the BB loop (purple) and in other loops (green). The site of the Lps^d mutation (Pro BB7) is coloured in

red. **b**, Drawing of the residues contributing to the conserved surface patch shown in **a**, in a similar orientation and colouring scheme. Also shown is the ion-pair between Arg BB3 and Glu α 13. **a** produced with Grasp²⁹ and **b** with Ribbons²⁸.

total of nine Toll homologues, but there is only one putative MyD88 homologue (CG2078). For this large number of receptors to signal through a common adapter molecule, the receptors must present a conserved surface area (S face) for coupling to the downstream adapter.

To identify the surface feature of the TIR domains that may be important for the recruitment of MyD88 (the S face), we analysed the molecular surface of the TIR domain in terms of amino-acid sequence conservation. This analysis reveals a large conserved surface patch (Fig. 2a), which mostly contains the BB loop, with additional contributions from helix α A, strand β B and the aromatic side chain of the (F/Y)DA motif (the β A-2 residue). The BB loop extends away for the rest of the TIR domain, forming a protrusion on the surface of the structure (Fig. 1a). This loop contains three highly conserved residues, Arg BB3, Asp BB4 and Gly BB8, in the RD Φ Φ Φ G motif (where Φ represents a hydrophobic residue, and x any residue). The side chains of both hydrophobic residues are exposed in the structure (Fig. 2a). The Arg BB3 residue has ion-pair interactions with the Asp BB4 residue (asparagine in TLR1 and TLR6 only) and the strictly conserved glutamate residue near the end of helix α A (Glu α 13) (Fig. 2b). The conformational difference for the BB loop in TLR2 (Fig. 1b) is probably due to the cacodylate modification of Cys β B4 during crystallization (see Methods), which is situated directly below the loop.

We performed functional studies to show that this conserved surface patch is crucial for signal transduction by the receptors. On the basis of the structural information, we mutated residues in this surface region and tested their functional effects using transfection assays with human TLR4 and *Drosophila* Toll^{10,15}. These studies show that mutations at almost every position in this region lead to significantly reduced signalling activity of the receptors (Fig. 3a). Residues that were mutated include BB3, BB4, BB7 and BB8 in the BB loop, and residue α A13 (Fig. 3a). In addition, it has been reported that mutation of the BB3 residue in IL-1RI (R to A)¹⁶ and the β A-2 residue of *Drosophila* Toll (F to I)⁵ also blocked receptor signalling. The ion-pair between Arg BB3 and Glu α A13 seems to stabilize the conformation of the BB loop. Notably, the double mutant, BB3 R to E and α A13 E to R, could not rescue the signalling activity (Fig. 3a). It is likely that this ion-pair is involved in additional interactions in the signalling complex, or that the double mutant did not fully restore the ion-pair interactions.

The importance of this surface patch for receptor signalling is also emphasized by the fact that the Lps^d mutation⁶ in murine TLR4, P712H, is at the Φ ₂ position in the BB loop (residue BB7). To understand the structural basis for the elimination of receptor signalling caused by this mutation^{6,7}, we determined the crystal structure of the P681H mutant of the TLR2 TIR domain (Table 1). There are no significant structural differences between this mutant and the wild type (Fig. 1c). The mutation site is located at the tip of the BB loop, farthest from the rest of the TIR domain. The structures show that the proline residue does not have a special structural role. The lack of signalling by the Lps^d mutation is thus not due to disruption of the TIR domain structure itself, but rather to the disruption of a direct point of contact with other molecule(s), and specifically other TIR domains. Our observations are in contrast to those for the Fas death domain, where a naturally occurring mutation (V to N) eliminates signalling by destroying the native conformation of the domain¹⁷. The structural analysis is consistent with sequence and functional observations, as small hydrophobic residues (alanine, valine, isoleucine) are present at the BB7 position in several TIR domains (Fig. 1d). *Drosophila* Toll contains a valine residue at the BB7 position, and our studies showed that a V to P mutation has little effect on the function of the receptor (Fig. 3a). In contrast, a V to H mutation, equivalent to the Lps^d mutation of TLR4, as well as mutation to arginine or glutamate at the BB7 position, greatly reduced receptor function (Fig. 3a).

To understand further the molecular mechanism for the disruption of signalling by the Lps^d mutation and other changes of this conserved surface patch, we performed protein-binding assays between purified recombinant glutathione S-transferase (GST) fusion TIR domains and *in vitro* translated TIR domains. The binding assays show that there is a significant interaction between the TIR domains of MyD88 and TLR2, and also that the P681H

Table 1 Summary of crystal structure information

Structure	TLR1	TLR2	TLR2 (P681H)
X-ray source	CHESS A1	APS ComCAT	APS ComCAT
Maximum resolution (Å)	2.9	3.0	2.8
Cell parameters (a, c) (Å)	101.3, 137.9	121.2, 91.6	122.1, 92.1
R/free R factor* (%)	25.4/29.6	24.4/27.4	24.2/28.4

*R = $\sum_i |F_o| - F_c| / \sum_i F_o$

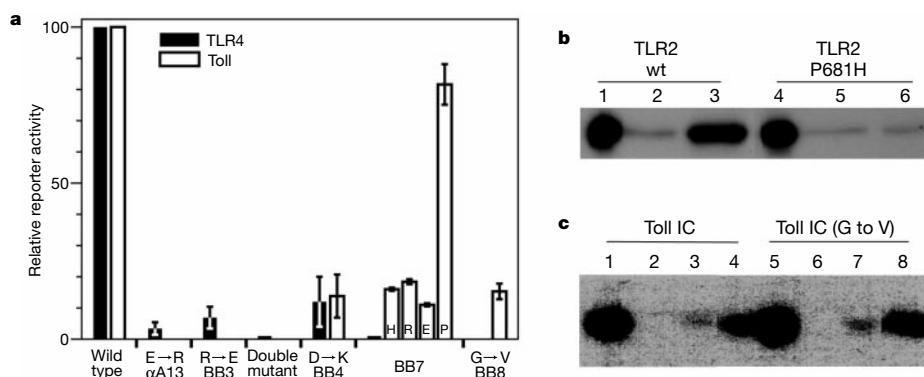


Figure 3 Functional studies of TIR domains. **a**, The relative signalling activity of mutants of human TLR4 (filled bars) and *Drosophila* Toll (open bars). At the BB7 position, the TLR4 mutant contained the P to H change (the Lps^d mutation), and the Toll mutant contains the V to H, R, E and P changes (as labelled). **b**, Protein-binding assays with MyD88 and TLR2 TIR domains, showing that the P681H mutation in TLR2 essentially abolished interactions

with MyD88. Lanes 1 and 4, 10% of input ³⁵S-labelled protein (for expression level control); 2 and 5, GST only (control); 3 and 6, GST+MyD88. **c**, Protein-binding assays with *Drosophila* Toll¹⁸, showing that the G→V mutation at the BB8 position does not interfere with receptor self-association. Lanes 1 and 5, 20% of input ³⁵S-labelled protein; 2 and 6, GST only (control); 3 and 7, GST+Toll IC; 4 and 8, GST+Pelle (K240R).

mutant of TLR2 almost completely abolished this interaction (Fig. 3b). At the same time, mutations in this surface patch do not seem to affect the oligomerization of the receptor (interactions at the R face), as shown by the binding assays with the G to V mutant at the BB8 position of *Drosophila* Toll (Fig. 3c). The experiment also shows that the previously documented interaction with the Pelle kinase was not affected by the mutation (Fig. 3c)¹⁸. Our biochemical data are supported by the observation that the intracellular domain of TLR4 containing the Lps^d mutation is a potent inhibitor of tumour-necrosis factor production in response to LPS stimulation as signalled by endogenous (full-length) TLR4 (ref. 19). Overall, the experimental data provide strong evidence that the R face is not disturbed by the Lps^d mutation. Both structural and functional studies therefore suggest that the conserved surface patch corresponds to the S face and that the Lps^d mutation abolishes receptor signalling by disrupting the recruitment of MyD88.

The backbone topology of the TIR domain is similar to that of the chemotaxis regulator CheY, as was proposed earlier²⁰, as well as many other proteins. The structural similarity between the TIR domain and CheY is however limited to the strands of the β -sheet (see Supplementary Information). In addition, TIR domains do not contain a cation-binding site equivalent to that in CheY (ref. 21), and therefore have a different mechanism of action from the response regulators. Nonetheless, it is interesting to note that a domain of similar architecture is used for signal transduction in response to exogenous stimulus in bacteria and eukaryotes.

We performed gel-filtration and dynamic light-scattering experiments to characterize the oligomerization state of isolated TIR domains in solution. These studies indicate that the inherent affinity for self-association of the TIR domains is rather low (dissociation constant (K_d) in the millimolar range; data not shown). At high concentrations, dimers and tetramers of TIR domains are observed in solution. This low affinity is consistent with the observation that signal transduction through TIR domains requires receptor oligomerization, induced either by ligand binding or by overexpression^{7–11}. It is likely that avidity may be important in the assembly of the TIR domain signalling complex. Our structural and functional information provide a molecular basis for understanding the mechanism and complexity of signal transduction by the TIR domains. □

Methods

Protein production and crystallization

Details on the expression, purification and crystallization of the TIR domains of human TLR1 and TLR2 will be presented elsewhere. Briefly, the TIR domain of human TLR1 (residues 625–786, about 20 residues from the transmembrane region) was overexpressed

in *Escherichia coli* and purified with nickel-agarose, cation exchange and gel-filtration chromatography. We purified the TIR domain of TLR2 (residues 626–784; about 16 residues from the transmembrane region) with cation exchange and gel-filtration chromatography. The P681H mutant of this domain was produced with the QuickChange mutagenesis kit (Stratagene) and sequenced to confirm the presence of the mutation. We obtained crystals of the TIR domain of TLR1 at 21 °C by the hanging-drop vapour-diffusion method. The reservoir solution contained 100 mM Tris (pH 8.0), 1.2 M NaH₂PO₄/K₂HPO₄, 5 mM dithiothreitol (DTT) and 20% glycerol. The crystals belong to space group P6₂2₂. The TIR domain of TLR2 was crystallized at 4 °C. The reservoir solution contained 100 mM cacodylate (pH 6.8), 10% PEG 8,000, 20% (v/v) DMSO, 200 mM MgCl₂ and 5 mM DTT. The crystals belong to space group P6₂2₂.

Structure determination

We collected X-ray diffraction data using rotating anode and synchrotron radiation sources (32-ID beamline (ComCAT) at APS, X4A beamline at NSLS and A-1 beamline at CHESS). The diffraction images were processed and scaled with the HKL package²². For TLR1, phases were obtained from the selenomethionyl multiwavelength anomalous diffraction (MAD) method²³ and the multiple isomorphous replacement (MIR) method²⁴. The atomic model was built with the program O (ref. 25). We carried out the structure refinement with the program CNS²⁶. There is a disulphide bond between Cys 707 and its twofold symmetry mate, which was confirmed by non-reducing SDS-PAGE and MALDI-TOF analysis on the crystals. A comparison with the structure of the TIR domain of TLR2 shows that the disulphide bond did not introduce significant conformational changes in the TIR domain of TLR1.

We determined the initial structure of the TIR domain of TLR2 by the combined molecular replacement protocol as implemented in the Replace package²⁷, using the structure of TIR domain of TLR1 as the model. The cysteine side chains had been modified by the cacodylate buffer, and the anomalous diffraction contained contributions from both Se and As atoms. The acentric reflections were phased with the anomalous diffraction data, and the centric reflections were phased with the molecular replacement solution. See Supplementary Information for more details on the structure determination.

There is only one molecule of the TIR domain in the asymmetric unit for crystals of both TLR1 and TLR2. This gives rise to a solvent content of about 80% and a unit cell volume to mass ratio of 5.5 Å³ per dalton for these crystals.

Functional studies with human TLR4 and *Drosophila* Toll

A truncated version of human TLR4, containing residues 558–825 and lacking the leucine-rich-repeat domain, was used for the mammalian functional studies as it has a higher constitutive activity. The mutants were made by site-directed polymerase chain reaction (PCR) mutagenesis and cloned into the pFLAG CMV1 vector. The receptor activity was measured by its ability to activate an NF- κ B-dependent luciferase reporter after transient transfections of 2 μ g of DNA into 293T cells¹⁰. The T110b mutant of Toll, which contains a C to Y mutation in its extracellular domain⁵, was used for the functional studies in *Drosophila* cells as it has a higher constitutive activity¹⁵. The mutants were made by site-directed mutagenesis with the QuickChange kit (Stratagene). We assayed signalling activity by co-transfection of 2 μ g of Toll and 0.4 μ g of dorsal expression vectors and a dorsal-dependent chloramphenicol acetyl transferase (CAT) reporter into *Drosophila* Schneider cells¹⁵. In both systems, all mutants were sequenced to confirm the presence of the respective mutations, and the assays were repeated several times to ensure reproducibility. Equivalent expression levels of the wild-type and mutant proteins were verified by western blotting.

Protein binding assays

The TIR domain of MyD88 was purified as a GST fusion protein, immobilized with

glutathione-agarose and incubated with [³⁵S]methionine-labelled wild type and P681H mutant of the TIR domain of TLR2 that were obtained by *in vitro* translation (Promega TNT system). For *Drosophila* Toll, the purified intracellular domain (GST–Toll IC) and GST–Pelle (K240R mutant) were incubated with [³⁵S]methionine-labelled intracellular domains of wild-type Toll and the G to V mutant at the BB8 position¹⁸. After washing, the bound proteins were eluted and separated by SDS–PAGE. Gels were fixed and exposed to X-ray film.

Received 12 July; accepted 11 September 2000.

- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. Jr & Ezekowitz, R. A. B. Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318 (1999).
- Kopp, E. B. & Medzhitov, R. The Toll-receptor family and control of innate immunity. *Curr. Opin. Immunol.* **11**, 13–18 (1999).
- Anderson, K. V. Toll signaling pathways in the innate immune response. *Curr. Opin. Immunol.* **12**, 13–19 (2000).
- Gay, N. J. & Keith, F. J. *Drosophila* Toll and IL-1 receptor. *Nature* **351**, 355–356 (1991).
- Schneider, D. S., Hudson, K. L., Lin, T.-Y. & Anderson, K. V. Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo. *Genes Dev.* **5**, 797–807 (1991).
- Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
- Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
- Muzio, M., Ji, J., Feng, P. & Dixit, V. M. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* **278**, 1612–1615 (1997).
- Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. MyD88, an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* **7**, 837–847 (1997).
- Medzhitov, R. *et al.* MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
- Burns, K. *et al.* MyD88, an adapter protein involved in interleukin-1 signaling. *J. Biol. Chem.* **273**, 12203–12209 (1998).
- Eck, M. J., Schoelson, S. E. & Harrison, S. C. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362**, 87–91 (1993).
- Mitcham, J. L. *et al.* T1/ST2 signaling establishes it as a member of an expanding interleukin-1 receptor family. *J. Biol. Chem.* **271**, 5777–5783 (1996).
- Burns, K. *et al.* Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nature Cell Biol.* **2**, 346–351 (2000).
- Norris, J. L. & Manley, J. L. Selective nuclear transport of the *Drosophila* morphogen dorsal can be established by a signaling pathway involving the transmembrane protein Toll and protein kinase A. *Genes Dev.* **6**, 1654–1667 (1992).
- Slack, J. L. *et al.* Identification of two major sites in the type I interleukin-1 receptor cytoplasmic region responsible for coupling to pro-inflammatory signaling pathways. *J. Biol. Chem.* **275**, 4670–4678 (2000).
- Huang, B., Eberstadt, M., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature* **384**, 638–641 (1996).
- Shen, B. & Manley, J. L. Phosphorylation modulates direct interactions between the Toll receptor, Pelle kinase and Tube. *Development* **125**, 4719–4728 (1998).
- Du, X., Poltorak, A., Silva, M. & Beutler, B. Analysis of Tlr4-mediated LPS signal transduction in macrophages by mutational modification of the receptor. *Blood Cells Mol. Dis.* **25**, 328–338 (1999).
- Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
- Stock, J. B., Stock, A. M. & Mottonen, J. M. Signal transduction in bacteria. *Nature* **344**, 395–400 (1990).
- Otinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Furey, W. & Swaminathan, S. PHASES-95: A program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.* **277**, 590–620 (1997).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. *et al.* Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Tong, L. Combined molecular replacement. *Acta Crystallogr. A* **52**, 905–921 (1998).
- Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature. The atomic coordinates have been deposited in the protein Data Bank under accession numbers IFYU, IFYW, IFYX.

Acknowledgements

We thank K. D'Amico and S. Wasserman for setting up the beamline at the Advanced Photon Source (APS) (supported by the US Department of Energy), R. Abramowitz and C. Ogata for setting up the beamline at the National Synchrotron Light Source (NSLS), and the MacCHESS staff for setting up the beamline at CHESS. We thank G. Bhargava, L. Duan and G. Xu for technical help; R. Khatyat, G. Jogi, and Z. Yang for help with data collection at the synchrotron sources; H. Wu and W. Hendrickson for discussions; and Columbia University (L.T.) and an NIH grant (J.L.M.) for financial support.

Correspondence and requests for materials should be addressed to L.T. (e-mail: tong@combio.columbia.edu).

Structural basis for the activation of 20S proteasomes by 11S regulators

Frank G. Whitby^{*†}, Eugene I. Masters^{*†}, Larissa Kramer^{*}, J. Randolph Knowlton^{*}, Yi Yao[‡], Ching C. Wang[‡] & Christopher P. Hill^{*}

^{*} Biochemistry Department, University of Utah, 50 N Medical Drive, Salt Lake City, Utah 84132, USA

[‡] Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA

[†] These authors contributed equally to this work

Most of the non-lysosomal proteolysis that occurs in eukaryotic cells is performed by a nonspecific and abundant barrel-shaped complex called the 20S proteasome¹. Substrates access the active sites, which are sequestered in an internal chamber, by traversing a narrow opening² (α -annulus) that is blocked in the unliganded 20S proteasome by amino-terminal sequences of α -subunits³. Peptide products probably exit the 20S proteasome through the same opening. 11S regulators (also called PA26 (ref. 4), PA28 (ref. 5) and REG^{6,7}) are heptamers^{4,8,9} that stimulate 20S proteasome peptidase activity *in vitro* and may facilitate product release *in vivo*. Here we report the co-crystal structure of yeast 20S proteasome with the 11S regulator from *Trypanosoma brucei*⁴ (PA26). PA26 carboxy-terminal tails provide binding affinity by inserting into pockets on the 20S proteasome, and PA26 activation loops induce conformational changes in α -subunits that open the gate separating the proteasome interior from the intracellular environment. The reduction in processivity expected for an open conformation of the exit gate may explain the role of 11S regulators in the production of ligands for major histocompatibility complex class I molecules^{10,11}.

After unsuccessful efforts to crystallize a complex of an 11S regulator and a 20S proteasome from the same species, we attempted to crystallize heterologous complexes. This approach was justified by the ability of 11S regulators to activate 20S proteasomes from widely divergent species. For example, trypanosome PA26 activates 20S proteasome from rat⁴, and human REG α activates 20S proteasome from yeast (M. Rechsteiner, personal communication). We therefore used the fluorogenic substrate suc-LLVY-MCA⁴ to show that recombinant PA26 stimulates the peptidase activity of yeast 20S proteasome by up to eightfold (data not shown).

We co-crystallized trypanosome PA26 and yeast 20S proteasome, and determined the structure at 3.2 Å resolution using the known structure of yeast 20S proteasome³ as a molecular replacement search model (Fig. 1a). The asymmetric unit contains a complex of one 20S proteasome and two PA26 heptamers that has a relative molecular mass (M_r) of 1.1 million, a length of 290 Å and a maximum diameter of 120 Å (Fig. 1b). This structure resembles electron microscopy images of mammalian 11S regulator/20S proteasome complexes¹².

Despite sharing only 14% sequence identity, the structure of PA26 resembles that of the human 11S regulator, REG α ⁸. The monomer, which comprises four long helices, packs into a heptamer that has overall dimensions of 90 Å in diameter and 70 Å in length (Fig. 2). Helices 2, 3 and 4, and the loop between helices 2 and 3 are quite similar in PA26 and REG α with r.m.s. deviation of 1.8 Å and maximum displacement of 3.6 Å for overlap on all 924 pairs of C α atoms. Helix 1 occupies the same volume in PA26 and REG α , although it is not clear how these residues should be aligned, in part because Fourier maps are more poorly defined in this section. The heptamer surrounds a central pore of 33 Å in diameter at the proteasome-binding end and 24 Å in diameter at the distal end. The loop between helices 2 and 3 is primarily responsible for

proteasome activation¹³, and seven of these activation loops encircle the opening to the central pore on the proteasome-binding face of the PA26 heptamer (Fig. 2c). Neighbouring the activation loops, at a larger radius, are the C-terminal sequences of PA26, which have

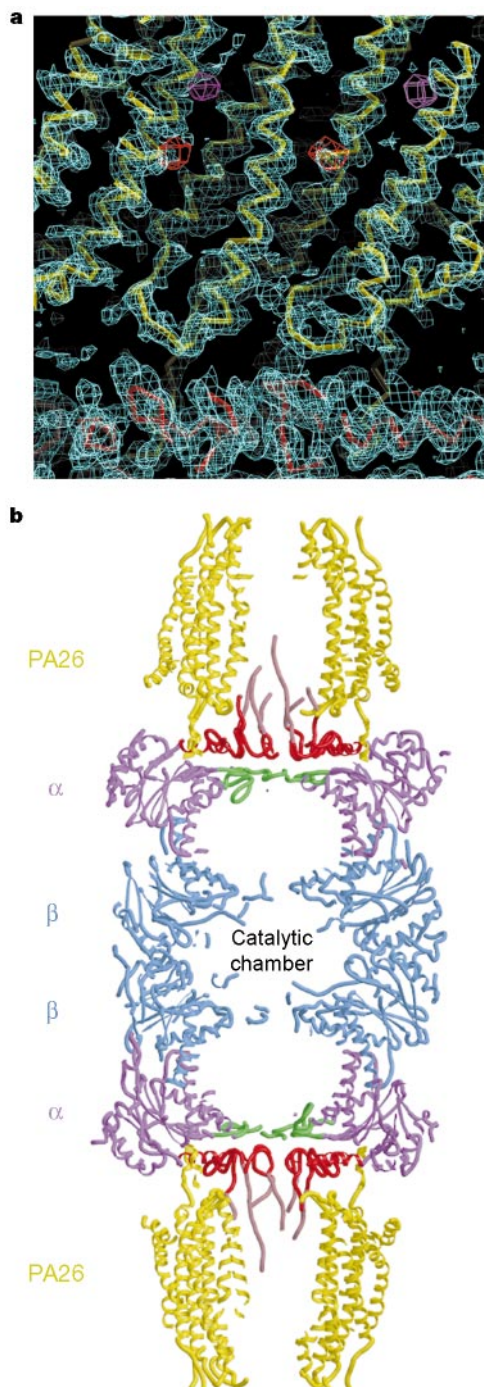


Figure 1 Structure of the PA26–proteasome complex. **a**, Unbiased density for PA26 contoured at 0.8 times the r.m.s. deviation. Phases were calculated from a 20S proteasome model and refined by averaging²⁹. PA26 models were not included in any previous refinement calculations. C α atoms are coloured yellow in PA26 and red in the proteasome. Isomorphous difference maps (5 times the r.m.s. deviation) for mercury (purple) and platinum (red) derivatives confirm the assignment of Cys 83 and Met 204. **b**, Ribbon representation of the complex. Image has been cut away to reveal internal features. Residues of the α -annulus are green; ordered N-terminal residues of α -subunits that do not have counterparts in β -subunits are red; and disordered N-terminal residues of α -subunits are pink.

been implicated in the binding of mammalian 11S regulators to 20S proteasome^{14,15}.

The yeast 20S proteasome is an assembly with $M_r \approx 750$ K of 28 homologous protein subunits that are designated α - or β -type on the basis of sequence similarity. It is arranged as four stacked rings: the two end rings are comprised of seven different α -subunits ($\alpha 1$ – $\alpha 7$) and the two middle rings are comprised of seven different β -subunits ($\beta 1$ – $\beta 7$). The different α - and β -subunits each have a unique position in the proteasome structure³. The 20S proteasome is therefore a barrel-shaped assembly that contains a roughly seven-fold axis along the length of the barrel, and one exact two-fold axis in a perpendicular direction. The active sites are located in a central chamber at the amino termini of three of the β -subunits.

The overall structure of the yeast 20S proteasome in complex with PA26 is similar to that of the unliganded proteasome³. The β -subunits are essentially unchanged on binding PA26, with least squares overlap giving an r.m.s. deviation of 0.7 Å for all 1,461 C α atoms in a ring of β -subunits. This absence of conformational change in the 20S proteasome β -subunits is notable in view of previously published observations, which suggested that binding of 11S regulators might alter the conformation at the catalytic centres^{7,16}. The earlier observations may instead result from the ability of 11S regulators to open the axial gate, although it is also possible that binding of PA26 induces small but significant conformational changes in the β -subunits that are not apparent at the current resolution. Alternatively, it is possible that conformational changes in the proteasome β -subunits only occur for cognate complexes, rather than for the mixed-species complex that we have crystallized.

Proteasome α - and β -subunits have similar three-dimensional structures, although α -subunits have an additional N-terminal extension of ~ 35 residues, which form a helix (helix 0) and a reverse turn that lie across the top surface of the proteasome³. The reverse turn is preceded by an N-terminal strand that adopts one of two conformations, either turning away from the proteasome ($\alpha 1$, $\alpha 6$, $\alpha 7$), or packing closely to form a gate that blocks the α -annulus ($\alpha 2$, $\alpha 3$, $\alpha 4$ and to a lesser extent $\alpha 5$). Thus, the tight seal provided by the closed gate conformation of the unliganded 20S proteasome is achieved by substantial departures from seven-fold symmetry in the N-terminal sequences of α -subunits (Fig. 3a). The conformations of three α -subunits ($\alpha 1$, $\alpha 6$, $\alpha 7$) are essentially unchanged in the PA26–20S proteasome complex, as are most residues in the other four α -subunits. In contrast, the N-terminal strands of subunits $\alpha 2$, $\alpha 3$, $\alpha 4$ and $\alpha 5$ rearrange to open the axial pore by adopting a conformation similar to that of $\alpha 1$, $\alpha 6$ and $\alpha 7$ (Fig. 3).

The PA26 pore aligns with the roughly seven-fold axis of the 20S proteasome, and the α -subunit N-terminal sequences extend into the PA26 pore where they pass close to the activation loop before density disappears for the more N-terminal residues (Fig. 1b). The

Table 1 Crystallographic statistics

Resolution range	40–3.2 Å	(3.26–3.2 Å)
Number of reflections measured	503,540	(16,790)
Number of unique reflections	189,615	(6,632)
Completeness	89.1%	(62.9%)
R_{merge}^*	0.112	(0.331)
$I/\sigma(I)$	7.0	(1.9)
$R_{\text{factor}}^\dagger$	0.255	(0.382)
$R_{\text{free}}^\ddagger$	0.325	(0.404)
r.m.s. bond lengths§	0.013 Å	
r.m.s. bond angles§	1.775°	
Average B: proteasome	72 Å ²	
Average B: PA26	89 Å ²	

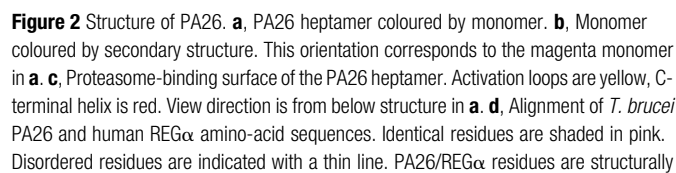
Values in parentheses refer to the highest resolution shell.

* $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the intensity of an individual measurement and $\langle I \rangle$ is the corresponding mean value.

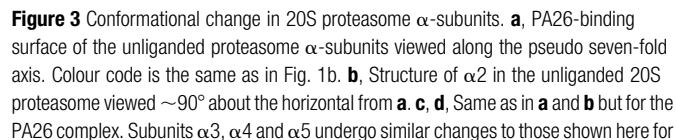
† $R_{\text{factor}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ is the observed and $|F_c|$ the calculated structure factor amplitude.

‡ R_{free} is the same as R_{factor} , calculated with a randomly selected test set of 1,028 reflections that were never used in refinement calculations.

§ The r.m.s. deviation in bond lengths and bond angles from ideal values.



equivalent from the start of helix 2 to the C terminus as indicated. The alignment for helix 1 is tentative. The hexahistidine affinity tag inserted after the initiator methionine of PA26 is shown. Every tenth residue is marked with a dash (the first PA26 residue marked is Gln10). Residues that contact the 20S proteasome are indicated with blue triangles. The PA26 construct used in this study has a threonine in place of the authentic serine⁴ at position 226.



$\alpha 2$, **e**, $F_o - F_c$ density (2.7 times the r.m.s. deviation) for α -subunit N-terminal residues. Map calculation followed torsion angle dynamics refinement in which these residues were deleted. C α coordinates are blue in the unliganded conformation and red in the PA26 complex. In each case, the view direction is equivalent to that shown in **b**.

residues that lack electron density presumably continue to extend into the PA26 pore but become flexible. Even when side chains are taken into account, this model can leave an opening that has a similar diameter to that of the α -annulus (13 Å).

The C-terminal residues of 11S regulators have a chief role in binding to the 20S proteasome^{14,15}. It is therefore notable that the different homologues, which are able to activate 20S proteasomes from disparate sources, have highly divergent C-terminal sequences. The C-terminal residues are inherently flexible⁸, but become ordered in the PA26–proteasome complex where they insert into pockets located between proteasome α -subunits (Fig. 4) that are preformed in the unliganded 20S proteasome structure³. The spacing of the seven proteasome pockets matches that of the seven PA26 C-terminal tails, thereby allowing the sequence-insensitive, and presumably weak, interactions of each individual C-terminal tail to function cooperatively.

The PA26 C-terminal tails are built with residues 224–225 in an extended conformation and the last six residues (226–231) as a

helix. Fourier maps are not well defined for the C-terminal residues. Five of the pockets contain density that matches a helical conformation for the C-terminal tail, whereas the other two pockets (between $\alpha 6$ and $\alpha 7$, and $\alpha 7$ and $\alpha 1$) seem to be empty. This variable degree of order presumably results from differences in the 20S proteasome α -subunit sequences, and it is possible that different 11S regulator tails bind preferentially to different subsets of pockets. Because of the marginal quality of the Fourier maps, we are not certain about details of the structure in this region; a helix is our best interpretation. An attractive feature of the helix model is that it places the C-terminal carboxylate in hydrogen-bonding distance of main-chain amide groups at the N terminus of helix 1 in 20S proteasome α -subunits. The hydrogen-bonding interaction between a charged carboxylate and a helix N terminus is expected to be relatively strong, and summing this contribution over several subunits may provide substantial binding affinity. This architecture explains how the highly variable C-terminal sequences are able to provide binding energy, and why deletion of the last C-terminal residue

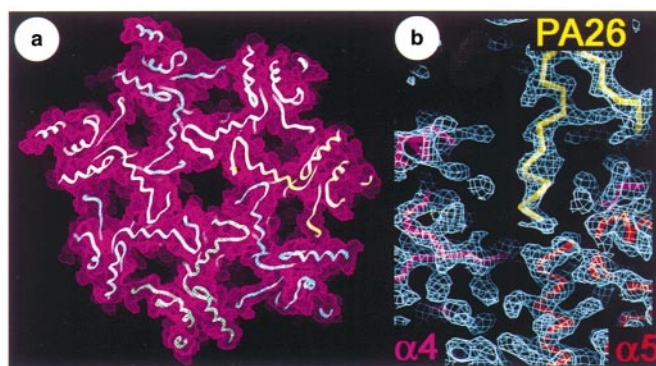


Figure 4 Mechanism of binding. **a**, Top view of the 20S proteasome α -subunits with the molecular surface shown as a purple net. The seven pockets that receive PA26 C termini surround the central open gate. **b**, Side view of PA26 C-terminal residues (yellow) binding into a pocket. This view is a close up of the interface at the top left of Fig. 1b. The PA26 C-terminal carboxylate approaches the N terminus of $\alpha 5$ helix 1, which is in a vertical

orientation in the lower part of this figure. $2F_o - F_c$ density (0.8 times the r.m.s. deviation) was computed following torsion angle dynamics refinement in which the last 30 residues of PA26 were set to zero occupancy. Density for the last well-defined side chain of PA26, Arg 223, appears to connect the two segments of PA26 near the top of this figure.

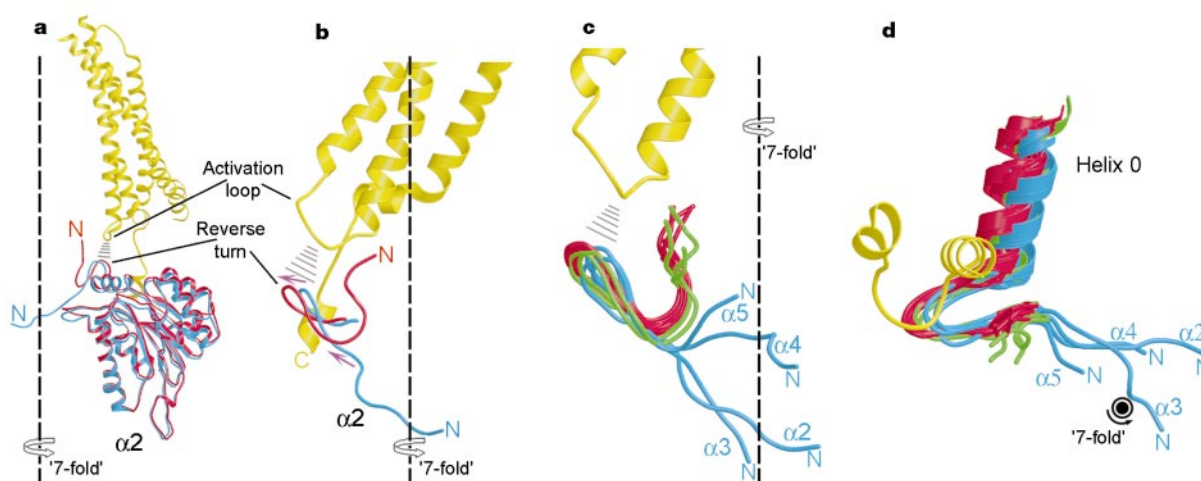


Figure 5 Mechanism of activation. **a**, Unliganded and complexed proteasomes were overlapped on their β -subunits. PA26 monomer (yellow); bound 20S proteasome $\alpha 2$ subunit (red); unliganded $\alpha 2$ conformation (blue). The PA26 C-terminal helix is visible behind $\alpha 2$, where it contacts groups near the N terminus of $\alpha 3$ helix 1 (not shown). Disordered N-terminal residues of α -subunits are not shown. View direction is similar to that of Fig. 3b. **b**, Close-up view, rotated by $\sim 90^\circ$ about the vertical axis in **a**. The $\alpha 2$ reverse turn and more N-terminal residues are displaced by about 2.5 Å (purple arrows). **c**, Reverse turn conformations for all α -subunits. $\alpha 1$ –7 of complexed proteasome (red);

$\alpha 2$ –5 of unliganded proteasome (blue); $\alpha 1$, 6 and 7 of unliganded proteasome (green). Overlap on proteasome β -subunits was followed by seven-fold rotations defined by the PA26 heptamer. $\alpha 2$ –5 all undergo a similar displacement of reverse turn and N-terminal residues. **d**, Same as in **c** but viewed along the seven-fold axis. The displacement of helix 0 visible in this figure results from close contact between helix 0 and the reverse turn of the neighbouring ($\alpha 1$) proteasome subunit, rather than from direct contact between helix 0 and PA26.

from REG α /PA28 α is more deleterious than its substitution¹⁴.

The activation loop seems to function as a rigidly constrained unit that induces the 20S proteasome to adopt an activated conformation. Binding affinity provided by C-terminal residues is used to open the 20S proteasome gate by pressing the activation loops, which have exactly seven-fold symmetry, against the pseudo seven-fold-symmetric N-terminal reverse turns of 20S proteasome α -subunits (Fig. 5). These contacts force the proteasome α -subunits to become more symmetric and therefore incompatible with the closed gate conformation. Overlap of the unliganded 20S proteasome³ and the PA26–20S proteasome complex shows that the reverse turns of subunits α_2 , α_3 , α_4 and α_5 each move by ~ 2.5 Å against the wedge-shaped activation loop. This motion propagates to an ~ 2.5 Å shift of more N-terminal residues away from the approximate seven-fold axis (Fig. 5b), thereby disrupting the numerous hydrogen bond and van der Waals interactions that stabilize the closed gate conformation. The N-terminal residues consequently rearrange to extend away from the proteasome surface and open the gate. The reverse turns and N-terminal extensions of proteasome α_1 , α_6 and α_7 subunits permanently reside in a conformation consistent with the open gate, and do not significantly change conformation on complex formation.

Unlike 11S regulators, which only activate the 20S proteasome against peptide substrates *in vitro*^{5,8}, the 19S regulatory complex delivers protein substrates for degradation by the 20S proteasome¹⁷. The observation that an 11S and a 19S regulator can bind simultaneously to each end of the same 20S proteasome¹⁸ suggests that the *in vivo* function of 11S regulators may be to open the product exit gate at the opposite end of the proteasome from the 19S regulator. This model suggests a function for 11S regulators in accelerating substrate throughput and also suggests a simple mechanism by which some of the mammalian 11S regulators may contribute to the production of peptides that are able to bind their histocompatibility complex (MHC) class I molecules^{10,11}. The requirement that peptides binding to MHC class I molecules have a hydrophobic or basic C-terminal side chain¹⁹ is met by inducible catalytic subunits that replace their constitutive counterparts to produce an 'immunoproteasome', which has reduced activity for cleavage after acidic side chains²⁰. The second requirement for binding to MHC class I molecules—that the peptide must be 8–10 residues in length¹⁹—presents a challenge, as proteasomes are highly processive and most of their products are shorter than eight residues²¹. Because product length is determined primarily by the ease of egress from the proteasome interior²², we propose that opening of the 20S proteasome exit gate by 11S regulators will reduce processivity and allow the release of longer peptides. These extended peptides can be trimmed from their N terminus²³, as necessary, to give the appropriate 8–10-residue length and hydrophobic/basic C-terminal residue. Regardless of the precise function in antigen production, the structure described here shows the general principles by which 11S regulators bind and activate 20S proteasomes. □

Methods

Protein purification

We prepared recombinant *T. brucei* PA26 by a modification of described methods⁴. *Escherichia coli* BL21(DE3) cells were transformed with pBtpa, which encodes the Val49 isoform of PA26 preceded by a hexahistidine affinity tag. Cells were grown, collected and the protein purified by Ni chelate chromatography. Fractions were identified by SDS–PAGE, pooled and dialysed against TSD200 (190 mM KCl, 10 mM NaCl, 1.1 mM MgCl₂, 0.1 mM EDTA and 10 mM Tris–HCl, pH 7.1) with 4 mM EDTA for 2 h; and then dialysed against TSD200 with 2 mM dithiothreitol for another 2 h. We concentrated and further purified the sample on a HiLoad 26/60 Superdex 200 Prep Grade gel-filtration column (Pharmacia). PA26 concentration was estimated by the Bradford method.

We prepared yeast 20S proteasome by a modification of described methods²⁴. Yeast cell strain SUB459, in which a His₆ affinity tag is extended to the C terminus of β_4 , was obtained as a kind gift from D. Finley (Harvard Medical School). Cells were grown to saturation in YM-1 media, collected and resuspended in buffer A (20 mM Tris–HCl pH 7.5, 1 mM EDTA), and then lysed in a beadbeater. Lysate was filtered through cheesecloth, centrifuged at 30,000g for 45 min, and ultracentrifuged at 142,000g for 60 min. Super-

natant was mixed with 300 ml of Fast Flow Q Sepharose resin (Pharmacia) pre-equilibrated in buffer A made up with 280 mM NaCl. Following packing into columns, protein was eluted with a linear gradient from 280 to 800 mM NaCl. Active fractions, assayed with suc-LLVY-MCA⁴, were dialysed against buffer B (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 15 mM imidazole, 5 mM MgCl₂), and applied to a Nickel–NTA–resin (Qiagen) column (1.0 cm interval diameter $\times$ 20 cm). After step elution by 100 mM imidazole in buffer B, we dialysed active fractions against buffer C (20 mM Tris–HCl pH 7.5, 1 mM EDTA, 200 mM NaCl), and concentrated and further purified them on a HiLoad 26/60 Superdex 200 Prep Grade gel-filtration column (Pharmacia). Fractions were pooled on the basis of activity and SDS–PAGE. Proteasome concentration was determined by ultraviolet absorbance assuming an extinction coefficient of 316,420 M cm⁻¹.

Crystallization and structure determination

PA26 and 20S proteasome were combined in a molar ratio of 2.5:1, dialysed against 10 mM Tris–HCl pH 7.5, 1 mM EDTA and 1 mM dithiothreitol, and concentrated to 10 mg ml⁻¹. Crystals were grown at 4 °C in sitting drops that contained 4 μ l of protein solution and 4 μ l of reservoir solution (0.1 M sodium HEPES, pH 7.5, 40% 2,4-methylpentanediol and 0.2 M NaCl). The space group was P2₁2₁2₁ ($a = 192.4$ Å, $b = 238.1$ Å, $c = 298.7$ Å) and the solvent content was 60%. Crystals were suspended in a nylon loop and cryocooled by plunging into liquid nitrogen before data collection at 100K. We collected data from a single crystal at beamline 9-1 of the Stanford Synchrotron Radiation Laboratory. After processing²⁵, most calculations used programs from the CCP4 suite²⁶. The known 20S proteasome structure³ (Protein Data Bank (PDB) 1ryp) was positioned by molecular replacement using AMoRe²⁷. An initial PA26 model was built into Fourier maps with the program O (ref. 28) and used to define operators for non-crystallographic (NCS) averaging with the program AVE (<http://alpha2.bmc.uu.se/~gerard/manuals/>). Rebuilding gave updated symmetry operators and molecular masks that were input to dm²⁹ for phase refinement by averaging and histogram shifting. Crystallographic refinement was performed using torsion-angle dynamics with XPLOR³⁰. NCS restraints were applied to 20S proteasome (2-fold) and PA26 (14-fold). Rebuilding was performed into maps calculated following torsion-angle dynamics refinement with the relevant section of the model set to zero occupancy. We applied a bulk solvent correction to the final refinement cycles and map calculations. Crystallographic statistics are given in Table 1.

Received 5 June; accepted 25 August 2000.

- Baumeister, W., Walz, J.K., Zuhl, F. & Seemuller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380 (1998).
- Wenzel, T. & Baumeister, W. Conformational constraints in protein degradation by the 20S proteasome. *Nature Struct. Biol.* **2**, 199–204 (1995).
- Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463–471 (1997).
- Yao, Y. *et al.* Structural and functional characterizations of the proteasome-activating protein PA26 from *Trypanosoma brucei*. *J. Biol. Chem.* **274**, 33921–33930 (1999).
- Ma, C.-P., Slaughter, C. A. & DeMartino, G. N. Identification, purification, and characterization of a protein activator (PA28) of the 20S proteasome (macropain). *J. Biol. Chem.* **267**, 10515–10523 (1992).
- Dubiel, W., Pratt, G., Ferrell, K. & Rechsteiner, M. Purification of an 11S regulator of the multicatalytic protease. *J. Biol. Chem.* **267**, 22369–22377 (1992).
- Realini, C. *et al.* Characterization of recombinant REG α , REG β and REG γ proteasome activators. *J. Biol. Chem.* **272**, 25483–25492 (1997).
- Knowlton, J. R. *et al.* Structure of the proteasome activator REG α (PA28 α). *Nature* **390**, 639–643 (1997).
- Zhang, Z. *et al.* Proteasome activator 11S REG or PA28: Recombinant REG α /REG β hetero-oligomers are heptamers. *Biochemistry* **38**, 5651–5658 (1999).
- Groettrup, M. *et al.* A role for the proteasome regulator PA28 α in antigen presentation. *Nature* **381**, 166–168 (1996).
- Preckel, T. *et al.* Impaired immunoproteasome assembly and immune responses in PA28^{-/-} mice. *Science* **286**, 2162–2165 (1999).
- Gray, C. W., Slaughter, C. A. & DeMartino, G. N. PA28 activator protein forms regulatory caps on proteasome stacked rings. *J. Mol. Biol.* **236**, 7–15 (1994).
- Zhang, Z. *et al.* Identification of an activation region in the proteasome activator REG α . *Proc. Natl Acad. Sci. USA* **95**, 2807–2811 (1998).
- Song, X., von Kampen, J., Slaughter, C. A. & DeMartino, G. N. Relative functions of the α and β subunits of the proteasome activator, PA28. *J. Biol. Chem.* **272**, 27994–28000 (1997).
- Li, J., Gao, X., Joss, L. & Rechsteiner, M. The proteasome activator 11 S REG or PA28: chimeras implicate carboxyl-terminal sequences in oligomerization and proteasome binding but not in the activation of specific proteasome catalytic subunits. *J. Mol. Biol.* **299**, 641–654 (2000).
- Dick, T. P. *et al.* Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* **86**, 253–262 (1996).
- Rechsteiner, M. In *Ubiquitin and the Biology of the Cell* (eds Peters, J.-M., Harris, J. R. & Finley, D.) 147–189 (Plenum, New York, 1998).
- Hendil, K. B., Khan, S. & Tanaka, K. Simultaneous binding of PA28 and PA700 activators to 20S proteasomes. *Biochem. J.* **332**, 749–754 (1998).
- Stern, L. J. & Wiley, D. C. Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure* **2**, 245–251 (1994).
- Gaczynska, M., Goldberg, A. L., Tanaka, K., Hendil, K. B. & Rock, K. L. Proteasome subunits X and Y alter peptidase activities in opposite ways to the interferon- γ -induced subunits LMP2 and LMP7. *J. Biol. Chem.* **271**, 17275–17280 (1996).
- Kisselev, A. F., Akopian, T. N., Woo, K. M. & Goldberg, A. L. The sizes of peptides generated from protein by mammalian 26 and 20S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem.* **274**, 3363–3371 (1999).
- Nussbaum, A. K. *et al.* Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl Acad. Sci. USA* **95**, 125024–125029 (1998).
- Craiu, A., Akopian, T., Goldberg, A. & Rock, K. L. Two distinct proteolytic processes in the generation

- of a major histocompatibility class I-presented peptide. *Proc. Natl Acad. Sci. USA* **94**, 10850–10855 (1997).
24. Rubin, D. M. *et al.* Identification of the *gal4* suppressor Sug1 as a subunit of the yeast 26S proteasome. *Nature* **379**, 655–657 (1996).
 25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
 26. Collaborative Computing Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 27. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
 28. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 29. Cowtan, K. & Main, P. Miscellaneous algorithms for density modification. *Acta Crystallogr. D* **54**, 487–493 (1998).
 30. Brünger, A. T. *X-PLOR Version 3.843, a System for X-ray Crystallography and NMR* (Yale Univ., New Haven, Connecticut, 1996).

Acknowledgements

We thank P. Brown, N.-L. Chan, B. Howard and M. Mathews for assistance with data collection; D. Finley for *Saccharomyces cerevisiae* strain SUB459, K. Alexander for assistance with protein purification; and W. Sundquist and M. Rechsteiner for comments on the manuscript. This work was supported by grants from the NIH and based on research conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the NIH (NCRR, NIGMS). Preliminary structure determination was performed with data collected from small crystals at Beamline X25 of the National Synchrotron Light Source, which is supported by the Office of Basic Energy Research of the US Department of Energy and by the NIH.

Correspondence and requests for materials should be addressed to C.P.H. (e-mail: chris@biochem.utah.edu). Coordinates are deposited in the PDB under accession number 1fnt.

erratum

Production of gene-targeted sheep by nuclear transfer from cultured somatic cells

K. J. McCreath, J. Howcroft, K. H. S. Campbell, A. Colman, A. E. Schnieke & A. J. Kind

Nature **405**, 1066–1069 (2000).

An incomplete version of Figure 1 was printed in this Letter. The correct Figure is reproduced below. □

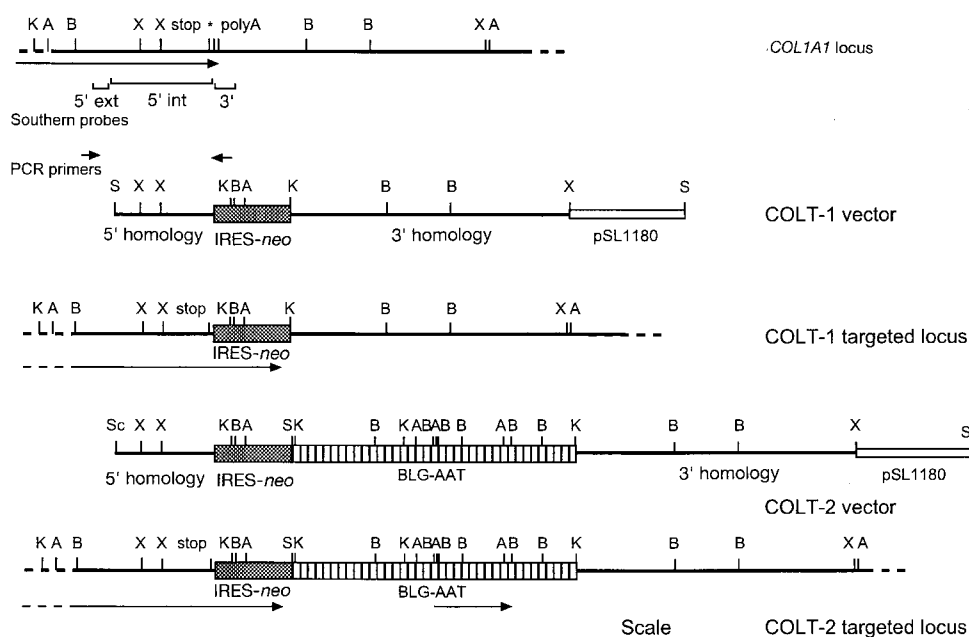


Figure 1 Diagrams of ovine *COL1A1*, COLT-1 and COLT-2 targeting vectors and targeted *COL1A1* locus. The map of ovine *COL1A1* shows the translational stop and polyadenylation sites, the direction of transcription and extent of the *COL1A1* mRNA is indicated by an arrow, an asterisk marks an *Ssp1* site where the targeted gene insertions were made. Southern hybridization probes and PCR primers are indicated. The maps of the COLT-1 and COLT-2 vectors and the targeted *COL1A1* locus show the IRES-*neo*

cassette as a shaded box, the BLG driven AAT transgene by a striped box, and the bacterial vector pSL1180 by an open box. The direction and predicted extent of the *COL1A1* / IRES-*neo* bicistronic mRNA and the AAT transgene mRNA are shown as arrows. The scale bar represents 2 kb. Restriction enzyme sites: A, *Asp1*; B, *BamHI*; Sc, *SacII*; S, *SalI*; K, *KpnI*; X, *XhoI*.